**UNITED STATES DISTRICT COURT**
**FOR THE SOUTHERN DISTRICT OF NEW YORK**

|  |  |
|---|---|
| LYNNE FREEMAN, <br><br> Plaintiff, <br><br> vs. <br><br> TRACY DEEBS-ELKENANEY P/K/A TRACY WOLFF, ENTANGLED PUBLISHING, LLC, HOLTZBRINCK PUBLISHERS, LLC D/B/A MACMILLAN, UNIVERSAL CITY STUDIOS LLC, EMILY SYLVAN KIM, AND PROSPECT AGENCY, LLC, <br><br> Defendants. | Case No.: 1:22-cv-02435-LLS-SN <br><br><br> **NOTICE OF REBUTTAL EXPERT DISCLOSURE** |

Defendants Tracy Deebs-Elkenaney p/k/a Tracy Wolff, Entangled Publishing, LLC, Holtzbrinck Publishers, LLC d/b/a MacMillan, and Universal City Studios LLC (collectively the "Entangled Defendants") hereby advise plaintiff Lynne Freeman that the Entangled Defendants intend to call Dr. Malcolm Coulthard as an expert rebuttal witness in the trial in this matter. Pursuant to Rule 26 of the Federal Rules of Civil Procedure, the Entangled Defendants are providing a rebuttal report from Dr. Coulthard which complies with Fed. R. Civ. P. 26(a)(2)(B). That report is attached hereto. Dr. Coulthard has an address at Flat 5, 90 Harborne Road, Birmingham, B15 3UH. The Entangled Defendants reserve the right to supplement this disclosure, and to name additional responsive or rebuttal expert witnesses, as necessary.

Dated: New York, New York
     August 31, 2023

COWAN, DEBAETS, ABRAHAMS,
& SHEPPARD, LLP

By: /s/ Benjamin S. Halperin
Nancy E. Wolff, Esq.
Benjamin S. Halperin, Esq.
CeCe M. Cole, Esq.

41 Madison Avenue, 38th Floor
New York, New York 10010
Telephone: (212) 974-7474
Fax: (212) 974-8474

*Attorneys for Defendants Tracy Deebs-Elkenaney p/k/a Tracy Wolff, Entangled Publishing, LLC, Holtzbrinck Publishers, LLC d/b/a MacMillan, and Universal City Studios LLC*

CONFIDENTIAL LITIGATION MATERIAL SUBJECT TO PROTECTIVE ORDER

**Expert Report of Dr. Malcolm Coulthard**

*Freeman v. Deebs-Elkenaney et al.*, **No. 1:22-cv-02435-LLS-SN (S.D.N.Y.)**

**Table of Contents**

## I.   **Introduction**

1.  I was retained by the CDAS law firm to provide an expert report responding to the opinions of Plaintiff Lynne Freeman's experts Drs. Carole Chaski and Patrick Juola. My opinions on these issues, which I am prepared to testify about at trial, are set forth below. I reserve the right to supplement or amend these opinions at counsel's or the Court's request.

2.  The reports submitted by Drs. Chaski and Juola do not support the Plaintiff's claim that Tracy Wolff's Crave book series plagiarises Freeman's unpublished manuscripts *Blue Moon Rising/Masqued*.

3.  The analytical techniques Drs. Chaski and Juola attempted to deploy have not to my knowledge been used previously in Forensic Linguistics to attempt to detect plagiarism in long works of fiction like those at issue in this case. These techniques, such as looking for short sequences of words that are shared, have been used to examine relatively short (typically maximum 500 words) texts, such as University admissions personal statements, police reports of interviews, and sworn witness statements. Applying these techniques to long works of fiction can introduce significant confusion and error. For example, if two works share several 6- or 7-word sequences (as Drs. Chaski and Juola claim), the significance will be very different in two works only 500 words long compared with works hundreds of thousands of words long. While I do not agree that it is appropriate to apply these techniques to the novels at issue here, I will still examine Drs. Chaski and Juola's methods and explain why they are otherwise unsound.

4.  Dr. Chaski opines that there are three types of plagiarism: (a) "copy-and-paste" plagiarism; (b) "mosaic" plagiarism; and (c) "conceptual" plagiarism, which she attempts to prove through an analysis of "keyword clusters." Each of these analyses is unreliable, and none of them supports the Plaintiff's claim of plagiarism.

5.  Dr. Chaski claims to show copy-and-paste plagiarism through "700 examples" of common 6-word sequences between the Freeman and Wolff works. But this claim is based on a severe misunderstanding of the evidentiary significance of common 6-word sequences. Six-word sequences are not evidentially significant in plagiarism detection; they are in fact expected when one compares two texts, particularly considering that authors often use multi-word idioms. Dr. Chaski does not cite any source that attributes significance to 6-word sequences (including my own work, which she repeatedly claims to rely upon).

Instead, the most recent thinking in our field is that textual borrowing is suggested (but not proven) when one finds (a) a shared 10-word sequence that (b) has no other occurrences in a Google search. Drs. Chaski and Juola have not reported identifying **a single** shared 10-word sequence, and thus do not move beyond the first step. Finally, and in any event, I understand that Dr. Chaski's claim of copy-and-paste plagiarism has now been withdrawn, since she did not provide copies of her claimed "700 examples" when asked to produce them.

6. Dr. Chaski's second claim, of "mosaic" plagiarism, is no more reliable. This analysis was based not on Dr. Chaski's own data, but on a pre-analysis by the Plaintiff, and was not adapted correctly to support her claimed copy-and-paste evidence. Dr. Chaski provides only 10 actual examples of alleged "mosaic" plagiarism, which were taken directly from the Plaintiff's Complaint and are presented without greater context or even page citations. Such limited data does not reliably suggest plagiarism, especially where there is no evidence that longer passages were lifted.

7. Dr. Chaski's analysis of "conceptual" plagiarism, intended to show that Freeman and Wolff wrote about similar concepts, is unreliable for numerous reasons. This analysis is a non-starter to begin with because it contradicts the fundamental principle that as linguists, we cannot detect the theft of ideas, but only the theft of text. The analysis also has no evidentiary value for at least the following reasons:

   a. Firstly, Dr. Chaski's system of keyword analysis does not work with objectively identified keywords; rather, it is based on an unexplained data selection for which it is impossible to derive a justification. Dr. Chaski chose to work with data that had been subjectively pre-selected by the Plaintiff in her Complaint specifically to support her plagiarism argument. Crucially, some words that appear key to only Freeman's books (for example "Tarot," "priestess," and "vision") were omitted, thereby concealing differences between the works.

   b. Secondly, if anything, a re-analysis of some of the "keywords" Dr. Chaski considered highlights differences between the texts, not similarities. For example, the words "moon" and "wolf" appear far more frequently in *BMR/Masqued* than they do in *Crave*, while the words "(San) Diego," "dragon," and "vampire" appear much more frequently in *Crave*. This kind of keyword analysis gives a different

2

snapshot, which suggests the opposite of what the Plaintiff and Dr. Chaski are claiming.

c.  Thirdly, after selecting her "keywords" based on reviewing the Plaintiff's Complaint, Dr. Chaski then analysed the comparative usage of these words in a corpus of 10 claimed "baseline" novels. But like her keywords, these "baseline" novels were not selected by Dr. Chaski herself, let alone objectively and with explicitly stated criteria, and not directed specifically at addressing the issue she was investigating. Rather, the corpus was chosen by the Plaintiff (or her lawyers), with no justification, and as we will see below, with some worrying deficiencies.

d.  Fourthly, Dr. Chaski's version of keyword analysis appears completely novel— essentially an experiment. I have certainly not encountered it before. Instead of investigating how frequently her keywords appeared in the works at issue, Dr. Chaski compared "overlap rates" of "lexical clusters" featuring her keywords. The method used in Johnson (1997), which Dr. Chaski claims to be generally following, is entirely different and did not examine "lexical clusters." Dr. Chaski is not following an established procedure with this analysis, and I know of no precedent in applying this technique to plagiarism detection.

e.  Finally, as her input was the product of an untried analytic procedure applied to subjectively collected data, it was impossible for Dr. Chaski to undertake a reliable statistical analysis and the results of her ostensibly sophisticated statistical tests must be disregarded because the input data were worthless.

8.  There are many fundamental problems with Dr. Juola's report. The greatest is that he attempts to demonstrate plagiarism by citing only **three** shared sequences that are 6 or more words long and **four** even shorter phrases. This evidence is meaningless given that the Wolff books are collectively almost nine hundred thousand words long. In examining 7-word sequences as supposed evidence of copying, Dr. Juola chiefly relies on (and misrepresents) my own prior scholarship. In all components of his analysis, Dr. Juola fails to reliably refute independent creation of the Wolff novels, because he cannot show that any of the few phrases that he has identified are uncommon, let alone unique to the two authors. In fact, phrases like "air as I try to" and "that million-dollar smile of his" **are** frequent according to Google, which defeats any notion that Wolff needed to copy them from Freeman.

3

Ultimately, Dr. Juola admits that his analysis is limited and that "it is not unbelievable that the Crave series and BMR are independently written." I would go further and argue that Dr. Juola (like Dr. Chaski) has provided no linguistic evidence to suggest that the Wolff books **not** independently created.

9. Neither Dr. Chaski nor Dr. Juola offers a persuasive (or even reasonable) explanation for **why** Wolff would want or need to copy short phrases like "air as I try to" from Freeman. As linguists analyzing plagiarism, we typically see plagiarists incorporate and then modify quite long sequences of other-authored text (as I will demonstrate below), not short, isolated phrases. Incorporating long sequences saves a plagiarist time and effort and so plagiarised sequences congregate and the evidence is self-reinforcing. Neither Dr. Chaski nor Dr. Juola has produced a single example showing Wolff plagiarising in this manner (or at all).

10. Finally, while Drs. Chaski and Juola have not provided reliable evidence to show that Wolff copied from Freeman, there **is** reliable evidence that Freeman herself copied from others. Below, I have shown that Freeman lifted a long, 89-word passage from the text of a famous television series in her 2013 *Masqued* manuscript. In illustrating what real plagiarism looks like, I will show the kind of evidence that Drs. Chaski and Juola needed, but failed, to produce in order to support the claim that Wolff plagiarised the *BMR/Masqued* manuscripts.

## II.   Preliminary Items

### A.      Qualifications

11. I hold the following degrees:

    a.  BA in English, with First Class Honours, University of Sheffield (1964);

    b.  Postgraduate Certificate of Education, London Institute of Education (1965);

    c.  Diploma in Phonetics, University College, London (1966);

    d.  MA in General Linguistics, University College London (1967);

    e.  PhD in English Language and Linguistics, University of Birmingham (1970).

12. I have held the following academic positions:

    a.  Research Fellow in Arts, University of Birmingham (1967-70);

    b.  Lecturer in English Language, University of Birmingham (1970-84);

    c.  Senior Lecturer in English Language, University of Birmingham (1984-93); Professor of English Language and Linguistics, University of Birmingham (1993-2004);

      d.  Professor of Forensic Linguistics, Aston University (2004-11);

      e.  Visiting Professor, Federal University of Santa Catarina, Brazil, (2012-17);

      f.  Emeritus Professor of Forensic Linguistics, Aston University (2012-present);

      g.  Emeritus Professor of English Linguistics, University of Birmingham (2004-present).

13. A brief Forensic Linguistics CV is attached as Appendix A.

## B.      Publications

14. In my career I have published 34 books and special editions of journals and some 100 journal articles and book chapters. According to Google Scholar there have been 23,758 citations of my publications and, in terms of the objectively measured significance of my work, I have an "h-index" of 42 and an i10 index of 24; (an h-index of 40 is considered "outstanding" and an i-10 index of 25 is considered "excellent").[1]

15. I was in fact the lead author of many of the publications Drs. Chaski and Juola cite in their expert reports. Drs. Chaski and Juola say they have drawn on the following of my publications:

      a.  M. Coulthard & A. Johnson, *An Introduction to Forensic Linguistics: Language in Evidence,* New York: Routledge (2007);

      b.  M. Coulthard, A. Johnson, & D. Wright, *An Introduction to Forensic Linguistics: Language in Evidence*, Routledge (2d Ed. 2017);

      c.  M. Coulthard & A. Johnson, *The Routledge Handbook of Forensic Linguistics*, Routledge (2010);

      d.  M. Coulthard, *Authorship Identification, Idiolect, and Linguistic Uniqueness*, 25 Applied Linguistics 431 (2004);

      e.  Malcolm Coulthard, *On admissible linguistic evidence*, 21 J. Law and Policy, 441 (2013);

      f.  T. Tayebi, & M. Coulthard, *New Trends in Forensic Linguistics*, 9 Language and Law/Linguagem e Direito 1 (2022).

16. A complete list of my publications from the last 10 years is attached as Appendix B.

---

[1] See Hirsch, J E. *An index to quantify an individual's scientific research output*, Proceedings of the National Academy of Sciences of the United States of America vol. 102,46 (2005): 16569-72. doi:10.1073/pnas.0507655102.

### C.        Achievements in the Area of Forensic Linguistics

17. I have been integrally involved in the international development of the field of Forensic Linguistics:

    a.   I helped found the International Association of Forensic Linguists and was the first President;

    b.   I co-founded the discipline's two major journals, *International Journal of Forensic Linguistics,* and *Language and Law - Linguagem e Direito.*

    c.   I helped to set up the first two masters degrees in the field at the UK universities of Cardiff and Aston;

    d.   I founded the world's first academic Centre for Forensic Linguistics at Aston;

    e.   I am the author of the two major text books in the field: *An Introduction to Forensic Linguistics* and *The Routledge Handbook of Forensic Linguistics.*

18. Further details of my work in Forensic Linguistics follow:

    a.   My last position before retirement was at the Federal University of Santa Catarina, where I was employed primarily to teach the analysis of written and spoken text and forensic language analysis. I established Forensic Linguistics as an academic subject in Brazil and was elected Honorary President of the newly formed Association of Language and Law, (ALIDI);

    b.   I am Emeritus Professor at the University of Birmingham, where in the 1990s I developed Forensic Linguistics as an undergraduate subject and trained, as doctoral students 8 of the current acknowledged leaders of the discipline, including Grant, Johnson and Olsson cited in the Chaski and Juola reports;

    c.   In 2001 I was appointed Honorary Professor at the University of Cardiff, where I helped to set up the world's first Master's degree in Forensic Linguistics which is still in rude good health;

    d.   I am also Emeritus Professor at the University of Aston, where in 2008 I established the world's first Centre for Forensic Linguistics with two members of staff. It now teaches forensic linguistics at Undergraduate and Postgraduate levels and hosted the Biennial conference of the International Association for Forensic and Legal Linguistics twice, in 2011 and in 2021. Now renamed as the Aston Institute for Forensic Linguistics it has a permanent staff of eight and has recently won a £6.7

million, ($8 million), five-year government grant to research five separate areas of forensic linguistics and employ 11 post-doc researchers;

e. I have travelled widely to teach courses, lecture, present research papers, and examine in: Argentina, Australia, Belgium, Brazil, Chile, Croatia, Cyprus, Czech Republic, Denmark, Egypt, Finland, France, Germany, Greece, Holland, Hong Kong, Iran, Ireland, Israel, Italy, Malta, Japan, Malaysia, Poland, Portugal, Singapore, Spain, Sweden, Tunisia, Uruguay, the USA and Venezuela;

f. In 1993 I was the Chair of the Founding Committee of the International Association of Forensic Linguists (now renamed the International Association for Forensic and Legal Linguistics) and became its first President;

g. I was also the founding co-editor of the two main forensic linguistics journals: *International Journal of Forensic Linguistics* (1994-present) and *Language and Law - Linguagem e Direito* (from 2015-present);

h. I organised the First British Seminar on Forensic Linguistics in March 1992, the Second British Seminar on Forensic Linguistics in 1993 and the Fourth Biennial Conference of the International Association of Forensic Linguists in 1999, all of them hosted by the Centre for Advanced Research in English at Birmingham University, of which I was the Founding Director.  Since I went to Aston the group has hosted the biennial Conference twice, in 2011 and 2021, and while I was working in Brazil, we hosted international conferences in 2013 and 2018;

i. Since 2000, along with ex-student Dr. K. Kredens, I have run the annual International Summer School in Forensic Linguistic Analysis, which from 2006 to 2011 was hosted by Aston University, the most recent summer school was in Manila in July of this year.

### D.      Prior Testimony

19. I have been commissioned to prepare reports for both prosecution and defense in over 230 cases, including eight reports for the England and Wales Criminal Cases Review Commission, one for the Scottish Criminal Cases Review Commission, one for the Police Ombudsman in Northern Ireland and two for the financial detective companies Kroll and Control Risks. I wrote reports for the following UK appeals against conviction: the Birmingham Six, Derek Bentley, Paul Blackburn, the Bridgewater Four, Robert Brown, Reg

Dudley and Ian Maynard, Iain Hay Gordon, Paul Malone, Peter Tredget and the Campbell 'Glasgow Ice Cream Wars' Appeal.

20. I have given evidence in court many times, including a case of academic plagiarism in the High Court in Hong Kong, three terrorist trials in Northern Ireland, a murder trial in Norwich, the Danielle Jones/Stuart Campbell 'texting' murder trial in Chelmsford, a Court Martial in Germany and the Derek Bentley, Paul Malone, Robert Burton, Paul Blackburn and Peter Tredget Appeals in the Royal Courts of Justice in London. I have also worked with the Metropolitan, the Scottish, the South Wales and the Military police on internal investigations. My most recent court appearance was in the Peter Tredget Appeal in the Royal Courts of Justice in London, October 2021. My most recent completed authorship cases were the $50 billion (sic) Yukos Arbitration case, which was settled in the Hague in February 2020 and a successful Appeal in Brazil for which I produced evidence of plagiarism in the final Judgment.

21. Cases from the last four years:

   a. **Yukos v. The Russian Federation (2016-2019).** I produced a series of four reports, co-authored with my colleague Professor Dr. Tim Grant, challenging a series of expert reports written by Drs. Carole Chaski and Walter Daelemans, in which they attempted to assign authorship to sections of an Arbitration Judgment, and thereby to invalidate it. We demonstrated that they had failed to do so. We were not called to give evidence in person, but in February 2020, the Hague Court of Appeal, explicitly citing our report, found in favour of the original owners of Yukos, by whom we had been contracted, and ordered the Russian Federation to pay $50 billion (sic) in compensation. The Russian Federation immediately appealed to the Dutch Supreme Court and the case is still *sub judice.*

   b. **Paul Tredget Appeal (2018-2022).** The last of a series of appeals against conviction for arson, at the age of 17, on the grounds that the police evidence was tainted. I was able to demonstrate that some of the evidence had been 'created' by the interviewing officer by means of suggestion and reformulation. I wrote two reports and gave evidence in person in October 2021. On the basis of my evidence and some support from the opposing expert, Dr. Olsson, the judges reversed some, but not all, of the grounds for conviction, in February 2022.

c. **Immortals LLC v. Fernando Alvarenga F. Costa (2021-2023).** This was a claim for liquidated damages for breach of a non-disparagement clause. While the offence was constituted by things said in an online discussion program in Portuguese in Rio de Janeiro, Brazil, the company and lawyers were based in the United States and the claim was scheduled for arbitration in English in LA. I co-authored the report with Dr. Carmen Rosa Caldas-Coulthard, a Brazilian native speaker of Portuguese. We pointed out the naiveté of assuming an English translation could successfully carry the meanings necessary for arbitration in English, especially when the translation had been undertaken by a third-year undergraduate medical student. We pointed out mistakes in the translation, linguistic problems with the relevant disparagement clause in the contract and with the apparent intention to work from written transcripts and not from subtitled video-recordings. Our evidence was not tested in court as Immortals chose to settle out of court for a fraction of the damages claimed initially. The defence lawyer said our report was crucially important in the decision to settle.

d. **Appeal of ex-Governor Pezao (2021-2023).** This was an appeal against conviction for financial corruption. This report was also co-authored with Dr. Carmen Rosa Caldas-Coulthard, as again the crucial texts were all in Portuguese. The claim was that the judge in the case has not been impartial. The report we wrote supported this claim by showing that while (a) the judge did not engage at all with a defense-written argument, which was some 120 pages long (he included no quotations, nor even paraphrases, of the arguments if only to dismiss them); (b) he not only relied heavily on the Prosecution document that was not much longer, but did so secretly—that is he almost never indicated when he was using their arguments, not even when using their supportive evidence. The central section of the Judgment comprised 64 pages, 51 of which contained plagiarised text; indeed, some pages consisted almost entirely of plagiarised text. We were not called to give evidence and do not know what weight, if any, was given to our report by the three appeal court judges, but in a judgment published in April of this year the conviction was quashed.

E.        **Compensation**

22. I am being compensated for my work in this case at a rate of $450/hour, both for preparing this report and for any testimony that I may be subsequently required to provide.

**F.        Materials Considered**

23. I considered the following materials in formulating my opinions and preparing this report:

      a.   Dr. Chaski's "Expert Report Regarding Plagiarism of Freeman Manuscripts," dated May 10, 2023.

      b.   Dr. Juola's amended "Report on Probabilistic Analysis of Overlaps in Young Adult Paranormal Romance Fiction," dated May 12, 2023;

      c.   The 2011 draft of Freeman's manuscript entitled *Blue Moon Rising* with Bates beginning LF03750 (hereinafter the "2011 Draft" or "*BMR*");

      d.   The 2013 draft of Freeman's manuscript entitled *Masqued* with Bates beginning LF06904 (hereinafter the "2013 Draft" or "*Masqued*");

      e.   The books *Crave, Crush, Covet,* and *Court* by Tracy Wolff (focusing my own analysis on *Crave*);

      f.   Freeman's First Amended Complaint ("FAC" or "Complaint");

24. Additionally, I requested the following data which is discussed in, but not actually appended to, Dr. Chaski's report: (a) "lemmatized" versions of the works she had created; (b) the "700 examples of 6-words-in-a-row exact matches" she "considered"; and (c) her "key word clusters." Item (b) was not provided at all and I was informed that items (a) and (c) were provided by the Plaintiff's attorneys but only on the evening of August 23. I am advised that the materials are voluminous and I have not had time to review them but I reserve the right to issue a supplemental opinion regarding them if so requested and permitted.

**G.        Turnitin**

25. In preparing my brief illustrative analysis of self- and other-plagiarism in Freeman's 2013 Draft *Masqued*, I made use of the internationally recognised plagiarism-detection program Turnitin.[2] This program searches the internet for matches of phrases in the document being analysed. Turnitin is quite a crude tool, as it typically finds the most recently published examples of phrases and these, given the massive amount of reused text on the internet, are rarely the first instances. However, Turnitin does indicate that the researched phrase has been used before. Then, given the fact that plagiarists rarely borrow just a single phrase,

---

[2] https://www.turnitin.com.

initial diagnosis by Turnitin allows more sophisticated follow-up searches involving more of the surrounding text, which can be used with Google often to identify the original source.

### III.   An Introduction To, And Application Of Forensic Linguistics

26. Forensic linguists, and myself in particular, have been growingly involved with authorship attribution since the mid-1980s. Most of the early cases either involved supporting convicted criminals' appeals against conviction by discovering linguistic evidence that police officers had falsified confessions, or were cases of student plagiarism; more recently work has involved the authorship attribution of text messages which has been used as evidence in murder cases.

27. I will now re-use a quotation from myself which was used by Dr Juola in ¶ 12 of his report, in order to outline the basis for using linguistic evidence in authorship cases:

> The underlying linguistic theory is that all speaker/writers of a given language have their own personal form of that language, technically labelled an *idiolect*. A speaker/writer's idiolect will manifest itself in distinctive and cumulatively unique rule-governed choices for encoding meaning linguistically in the written and spoken communications they produce. For example, in the case of vocabulary, every speaker/writer has a very large learned and stored set of words built up over many years. Such sets may differ slightly or considerably from the word sets that all other speaker/writers have similarly built up, in terms both of stored individual items in their passive vocabulary and, more importantly, in terms of their preferences for selecting and then combining these individual items in the production of texts.[3]

28. People compose sequences of words using two strategies: (a) the "open choice principle," where words are selected one by one; and (b) the "idiom principle" where phrases are selected ready formed and slotted into the developing sequence, like for example "one-by-one" in the line above or "million-dollar smile," or indeed "San Diego" from the texts under consideration. Because most written and spoken sequences are produced using both principles, it will always be easy to find texts sharing short sequences of words; what is idiolectally distinctive is the way a given author combines the idioms with the open choice words and thereby generates a unique sequence.

---

[3] M. Coulthard, *On admissible linguistic evidence*, J. Law and Policy, XXI(2):441-466 (2013) (cited in Juola Rep. ¶ 12, n.4).

### A.    On Authorship

29. As this case revolves around a dispute about authorship, it is essential to have a set of neutral terms to describe the complexity of the compositional process. Love (2002)[4] suggests that the process of authorship can, usefully, be broken down into four primary functions:

   a. *Executive Authorship*—the actual function of composition in terms of crafting the message and choosing the final wording of the text under consideration. So, I, Malcolm Coulthard, am the executive author of this report.

   b. *Precursory Authorship*—where a text draws on other texts, (which may themselves, in turn, have drawn on other earlier texts), by quoting from them, borrowing text from them or summarizing them. In this case the authors of these original sources are normally credited with precursory authorship. So, for example, below I incorporate extracts from the reports written by Drs. Juola and Chaski and even from my own writings.

   c. *Revisionary Authorship*—which includes both large- and small-scale editing of the text. This can be done by the Executive Author, (I am currently revising my first draft of this section), but also by others. So, for example, the final texts of the dissertations of my masters and doctoral students are partly the product of my revisionary comments, additions, deletions, and re-orderings.

   d. *Declarative Authorship*—the declaration or signing of a piece as one's own production, which claim can, of course, be partially or completely untrue. For example, as Dr. Juola demonstrated in 2013, J.K. Rowling is the executive author of books bearing the declarative name of Robert Galbraith. In this case, Tracy Wolff is the declarative author of the Crave series. However, Drs. Chaski and Juola argue that, using the terms I have just introduced, the series includes unacknowledged precursory language authored by Lynne Freeman. I will evaluate and reject this claim and the arguments adduced to sustain it, in what follows.

30. Any given individual may fulfill all of the above authorial functions for any given text, but also any number of individuals may separately perform each function.

31. I will now illustrate complex authorship by focusing on the opening and the closing sections of Freeman's 2013 Draft, *Masqued.* Although it has a different title, this manuscript is in

---

[4] Love, H., *Attributing Authorship,* Cambridge, Cambridge University Press (2002).

fact a substantially revised and shortened version of Freeman's earlier 2011 Draft, *BMR*. In the terms introduced above, we can discover in the opening section of the 2013 Draft, in addition to (new) executive text, both precursory text which has been borrowed from the 2011 Draft, and evidence of revisionary authorship. Below is an account of how the opening of Chapter 1 of the 2011 Draft was revised to become the opening of the Prologue of the 2013 Draft.

a. The *Masqued* Prologue begins with a section of approximately 400 new words ending in: "The telltale tingling at the base of my skull is the only warning I have as my spirit catapults forward into the World of Dreams, the shell of my body crumpling to the floor." (2013 Draft pp. 5-6.)

b. What follows is a mixture of precursory text taken from the beginning of the 2011 Draft and new executive text. In the excerpt below, incorporated precursory text is indicated by **bold and black**; pre-existing text that was rejected by a revisionary author (presumably Freeman), is indicated by ~~crossing out and red,~~ and new text, which could have been added by the executive, or by a revisionary, author, is indicated in blue:

> **The moon is full overhead, pregnant with possibilities and none of them good.**
> **Stunned, I lie on my back in the snow, my eyes full of the night sky above, longing for release—struggling for a breath that won't come. I'm suffocating. My mouth opens and closes like a fish out of water; my diaphragm spasms as I struggle for air, my body thrashing. I can't get my lungs to work.** ~~I know this is because I've fallen, and that in a few moments my breath will come back, yet my panic is overwhelming~~.
> **When I'm finally able to suck in air, it's in great, gulping, ragged gasps that shatter the stillness of the forest around me. I know I shouldn't make any noise but my body is so starved for oxygen that I have no control. The moon is my focus, my anchor, as the spasms slowly subside.**
> ~~I'm alone with who-knows-what looking for me~~. Where am I, ~~How did this~~ What's **happen**ed? ~~How had I gotten myself into this predicament in the first place?~~
> **I can't remember.**[5]

---

32. In this opening section we see examples both of what Dr. Chaski calls a "copy-and-paste" strategy—large sections reused verbatim—and of her "mosaic" strategies ("substitution, insertion, deletion and permutation of the source material"). (See Chaski Rep. ¶ 4.)[6]

33. The ending of Freeman's 2013 Draft (*Masqued*) is also a combination of new executive text followed by precursory text. In the following excerpt from page 468 of *Masqued*, the plain text at the top is new executive text and the section in italics below is 89 words of precursory text:

> Ash comes to Anna and foreswears his vows in order to be with her. To help her.
>
> *Voiceover: The tough thing about following your heart is what people forget to mention, that sometimes your heart takes you to places you shouldn't be, places that are as scary as they are exciting and as dangerous as they are alluring, and sometimes your heart takes you to places that can never lead to a happy ending.*
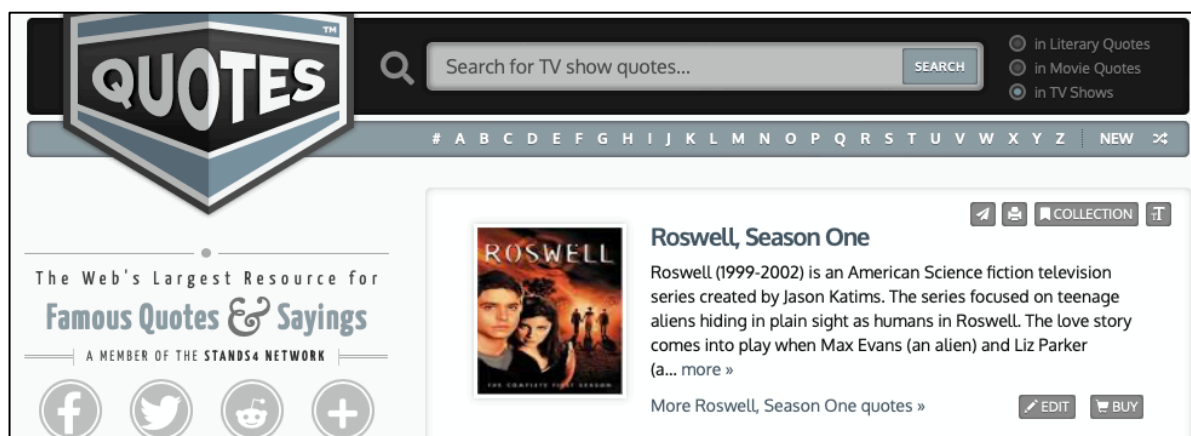> *And that's not even the difficult part. The difficult part is when you follow your heart, you leave normal, you go into the unknown...And once you do, you can never go back.*

34. One would imagine that the 89-word sequence of precursory text at the end of the above excerpt (the italics) also came from one of Freeman's prior drafts. However, that is actually not the case. Instead, the 89-word precursory text was taken verbatim by Freeman from somebody else's work. It turns out the precursory text comes from *Roswell*, an "American science fiction television series," which ran from 1999 to 2002, and which would be a natural source for Freeman to draw inspiration from.[7] The 89-word sequence is a direct quote from that TV show, and is available on several websites, one of which Freeman presumably accessed and copied from.[8]

35. For example, the following appears on the first of the websites just footnoted:

---

[6] While this strategy of incorporating earlier drafts (precursory text) without acknowledgement is standard practice for creative writers, academic institutions label the same process "self-plagiarism" because they are anxious that neither students nor staff members get credit twice for one piece of output.

[7] According to Wikipedia, "Roswell (1999–2002) is an American Science fiction television series created by Jason Katims. The series focused on teenage aliens hiding in plain sight as humans in Roswell. The love story comes into play when Max Evans (an alien) and Liz Parker (a human) fall in love. The series pilot was based on the Roswell High young adult book series, written by Laura J. Burns and Melinda Metz and published by Pocket Books." https://en.wikiquote.org/wiki/Roswell_(TV_series).

[8] See https://www.quotes.net/show-quote/68131; https://www.roswelloracle.com/journal.html.

36. I know that the two Freeman manuscripts I was asked to concentrate on are unpublished drafts and one assumes that the long sequence of copy-paste plagiarism just revealed would have been modified in a subsequent editing and publication process. But this remains a clear example of the sort of "copy-and-paste" plagiarism that Drs. Chaski and Juola have signally failed to discover in the Crave series.

37. Had *Crave* included such an 89-word sequence of verbatim precursory text from *BMR/Masqued,* this would have been clear evidence that whoever was responsible for inserting it had borrowed from a Freeman manuscript. However, Drs. Chaski and Juola have apparently found nothing even remotely comparable. **No such examples are provided in their reports.**

### B.    On Linguistic Plagiarism

38. Both Drs. Chaski and Juola talk about "plagiarism" but do not define it, let alone discuss how one can best identify it in a suspect text. So, it would be useful to do so here. Often "intention" is invoked and there are many definitions online, such as "Plagiarism can take

many forms, from **deliberate** cheating to accidentally copying from a source without acknowledgement."[9]

39. However, as linguists we cannot detect intention. As long ago as 1946, literary critics labeled such attempts to access intention as the "Intentional Fallacy."[10] The point was that one can only analyse the text, not derive the author's intention from it. So, while a linguist can identify unattributed precursory text, they cannot decide whether it is the result of "**deliberate** cheating [or] accidental copying." So, for our purposes let us define plagiarism as "the unacknowledged use of precursory text." Then, we must find a way to reliably detect precursory text.

40. Identifying **possible** precursory text is easy using any compare-text program, even Microsoft Word. Indeed, without saying how she did it, Dr. Chaski claims to have "considered over 700 examples of 6-words-in-a-row exact matches between the Tracy Wolff ("TW") documents and the Lynne Freeman ("LF") manuscripts." (Chaski Rep. ¶¶ 3, 61.) But finding the same phrase in two different documents does not mean that one author has copied from the other. The shared phrase could be a well-used idiom, or both authors could have created the phrase word-by-word independently, or one or both could have copied the phrase from a third author. So, we now ask: what more is needed to argue that finding pieces of text in a Wolff book that also occur in a Freeman manuscript must in itself be evidence that Wolff has incorporated precursory Freeman text?

41. The latest formulation of my thoughts on plagiarism detection is to be found in a forthcoming book chapter that I co-authored (which is attached as Appendix C). The formulation reads, with added bold:

> Coulthard (2004) argued that anyone generating a new meaningful, grammatically organised, sequence of words quickly produces a sequence that is unique, that is, they generate a sequence that no one, not even the author themself has ever produced before. Further empirical research over 20 years allows us to say that **if two texts share an identical unattributed 10-word sequence, for which Google can find no other instances,** there is **a possibility** that one author copied from the other or that both copied from a third text. If the identical sequence is even longer and especially

---

[9] https://www.student.unsw.edu.au/what-plagiarism.
[10] W. K. Wimsatt Jr. & M. C. Beardsley, *The Intentional Fallacy*, The Sewanee Review, Vol. 54, No. 3 (Jul.-Sept., 1946), pp. 468-488.

> if there are also several other such shared long identical sequences, and/or other almost identical shorter sequences linked by deletion, addition or paraphrase, the evidence for plagiarism can become persuasive.[11]

42. To rephrase, to reliably identify potential plagiarism, based on the occurrence of common text, empirical evidence suggests that it is important to first identify a sequence that is minimally 10-words long and for which Google can find ideally no other instances. The investigator must then support this finding with other examples of long sequences and/or reliable instances of what Dr. Chaski calls mosaic plagiarism. But even then, this only identifies possible plagiarism. The final confirmation is when the source and suspect texts are the only ones to share the extract. So, this is a three-step process:

   a. Step 1: identify a shared sequence at least 10 words long;

   b. Step 2: discover the rarity of this sequence using Google;

   c. Step 3: check if there are other linked copy-and-paste and/or mosaic sequences to support the plagiarism hypothesis.

43. As shown below, neither Dr. Chaski nor Dr. Juola has reliably completed any one of those steps.

## IV.   Evaluation Of Dr. Chaski's Expert Report/Opinions

44. In her report, Dr. Chaski proposes three analytical approaches to generate evidence to support the claim that Wolff plagiarised Freeman's drafts: "copy-paste, mosaic and conceptual plagiarism." (Chaski Rep. ¶ 2.) She claims that "using standard methods for detecting plagiarism in forensic linguistics, [she] found evidence of each type of plagiarism in the Crave series relating these books to the Freeman Manuscripts." (*Id.*) Each of these approaches is problematic. Let us analyse the three approaches one-by-one:

### A.       Copy-Paste Plagiarism

45. Dr. Chaski claims to have "reviewed over 700 examples of 6-words-in-a-row exact matches between the Tracy Wolff ("TW") documents and the Lynne Freeman ("LF") manuscripts." (Chaski Rep. ¶ 61.) The phrasing here is odd in that Dr. Chaski does not actually claim to have identified or collected the examples herself. As Dr. Chaski certainly accepted the corpus of comparator novels ready-selected by the Plaintiff (see my ¶ 73 below), it raises

---

[11] Coulthard, Frohlich & Jorge, *Forensic Linguistics in Brazil: Research and Practice*, p. 15, forthcoming in *Romance Forensic Linguistics*, Springer, (eds. Dieter Stein and Victoria Guillen Neto) (emphasis mine).

the possibility that this corpus of examples was also provided to Dr. Chaski by the Plaintiff or her lawyers. However, irrespective of how this corpus was assembled, these "700 examples"—which were never actually provided although I requested them—could not be reliable evidence of plagiarism for multiple reasons.

46. Firstly, Dr. Chaski asserts, claiming to draw on the research of others, that "since the consensus in forensic linguistics is that 6-words-in-a-row indicates textual borrowing, these 700 examples are evidence of copy-paste plagiarism." (Chaski Rep. ¶¶ 3, 59.) There are two crucial basic errors in this statement that invalidate Dr. Chaski's entire analysis:

   a. Contrary to what Dr. Chaski claims, there is in fact no "consensus in forensic linguistics that 6-words-in-a-row indicates textual borrowing"; and

   b. The existence of shared sequences of any length is not in itself evidence that one author has borrowed from the other or indeed has borrowed from anyone. As explained throughout this report, a shared sequence without corresponding reliable evidence of uniqueness simply does not in itself suggest, let alone prove, plagiarism.

47. Dr Chaski refers to a single source—Olsson (2008)[12]—to support of her claim of consensus although she does not quote from him. (Chaski Rep. ¶ 59 n.19.) But, in fact, the 2008 edition of Olsson's book is silent on the issue of what sequence length is needed to detect possible textual borrowing.

48. Dr. Chaski claims additional support, although she does not quote any, by stating "See also references in footnote 1." (Chaski Rep. ¶ 59 n.19 (referencing n.1).) But again, the sources listed there do not support her claim of a consensus:

   a. In his earlier, 2004 edition, mentioned once by Chaski, Olsson gives examples of only 7-, 8- and 9-word sequences, and even then, in his view, their evidentiary value needs to be reinforced by confirmation from Google that they are unique to the texts from which they have been taken. Notably, it was the same claim in my 2004 article *Authorship Identification, Idiolect, and Linguistic Uniqueness*,[13] to which Dr. Juola refers in his report. (Juola Rep. ¶ 14 n.7.) In that article, I demonstrated that the

---

[12] Olsson, John, *Forensic Linguistics*, Second Edition (2008).
[13] M. Coulthard, *Authorship Identification, Idiolect, and Linguistic Uniqueness*, Applied Linguistics, 25.4, 431-47 (2004).

police had falsified evidence. I showed, firstly, that two supposedly independently produced documents included identical sequences **and then** that these sequences were unique, because the shared examples did not occur at all in a Google search.

    b. Dr. Chaski's footnote no. 1 also cites my 2007 book written with Dr. Alison Johnson, *An Introduction to Forensic Linguistics: Language in Evidence*[14] and Dr. Johnson's 1997 article *Textual Kidnapping*.[15] These also do not support the claimed consensus—never have we argued, and certainly not in that book, that 6-word sequences are evidence of plagiarism, nor claimed that any particular sequence length is, in itself, sufficient.

49. To summarise, none of the sources Dr. Chaski cites in fact supports her claim that there is a "consensus in forensic linguistics that 6-words-in-a-row indicates textual borrowing." There is no such consensus. To the contrary, the most recent thinking is that sequences of a minimum of 10-words in length, **plus** evidence of uniqueness, are needed to **suggest** that there may have been textual borrowing. And of course, Dr. Chaski does not actually provide her 700 6-words-in-a-row examples, let alone show that they are unique to the Freeman and Wolff's texts.

50. If we apply the 10-word rule as an entry level search criterion for identifying precursory text (whether acknowledged or not by the author) to the books in this case, the results are persuasive. Neither Dr. Chaski nor Dr. Juola has identified **a single shared 10-word sequence** in the Freeman/Wolff works. In her analysis, Dr. Chaski (and Dr. Juola as explained below) mistakenly set the sequence length far too short, and also did not provide reliable evidence of uniqueness. Their derived results therefore provide no support for the Plaintiff's claim of plagiarism.

**B.**        **Mosaic Plagiarism**

51. Dr. Chaski's second analytic approach is to identify "mosaic" plagiarism.[16] Dr. Chaski observes that this type "is more difficult to detect than copy-paste, but it is detected through **paraphrase recognition, i.e., substitution, insertion, deletion and permutation of the source material."** (Chaski Rep. ¶ 4 (bold in the original).) I agree that it is much harder to

---

[14] M. Coulthard & A. Johnson., *An Introduction to Forensic Linguistics: Language in Evidence* (2007).

[15] A. Johnson, *Textual kidnapping: a case of plagiarism among three student texts*, in *Forensic Linguistics* (1997).

[16] Other authors have used the term "patchwriting," which seems to better capture the process of incorporating precursory text.

identify mosaic than copy-and-paste plagiarism, and indeed, Dr. Chaski does not tell us how she did it. Perhaps she didn't do it; perhaps she simply accepted examples that the Plaintiff (or her lawyers) had generated.

52. Surprisingly, Dr. Chaski begins by analysing the chapter titles in *Crave* and claiming that they show the use of "allusion." (Chaski Rep. ¶¶ 70-76 & tbls. 2A, 2B, 3.) As linguists, we usually call this *intertextuality*—the writer sets out to link their piece of text to an earlier well-known textual item. However, while using intertextuality is perhaps evidence of clever writing, it is not evidence of plagiarism. It is in fact astonishing that Dr. Chaski chose to treat a series of chapter titles in *Crave* (such as "We Came, We Fought, I Froze," "Is That a Wooden Stake in Your Pocket or Are You Just Happy to See Me?" and "Carpe Kill-Em") as if they were somehow examples of mosaic plagiarism, when they are just intertextual references to well-known phrases (in this case, as Dr. Chaski acknowledges in her table, to "I came, I fought, I conquered," "Is that a pistol in your pocket or are you just glad to see me?," and "Carpe diem," respectively). (Chaski Rep. tbl. 2A.) Clearly, Freeman is not the precursory author of the common phrases identified in Dr. Chaski's tables, and Dr. Chaski acknowledges that "no one disputes" the *Crave* chapter titles "are original to the TW books." (Chaski Rep. ¶ 76.)

53. Furthermore, for intertextuality to correctly function, the reader must be able to **recognise** the source, in order to appreciate the reference. So, this exercise does not advance the plagiarism argument, because a plagiarist is trying to **hide** the original source.

54. Dr. Chaski then provides ten specific examples of supposed mosaic plagiarism. (Chaski Rep. ¶¶ 78-87.) It appears that Dr. Chaski did not find these examples herself either, but rather merely copied them from the Plaintiff's documents. Certainly, the first nine of Dr. Chaski's examples (Chaski Rep. ¶¶ 77-86) were taken verbatim from the Plaintiff's Complaint. (FAC ¶¶ 41(a)(5), 42 (a)(1), 42(a)(7), 42(a)(9), 42(a)(11), 42(a)(18), 41(b)(8), 44(a)(36), 44(b)(8).) While Dr. Chaski claims to be aware of "many more examples" (Chaski Rep. ¶ 88, referencing the FAC), she does not list them for checking purposes nor show that any of them was identified independently and if so, using what methodology.

55. Isolated examples of claimed mosaic plagiarism such as those provided in ¶¶ 78-87 have no evidentiary value. The examples are presented without their context (or even sources) and have been altered through the use of ellipses and brackets.

56. There is also no underlying understanding of the nature of plagiarism in Dr. Chaski's treatment of mosaic plagiarism. Obviously mosaic plagiarism, rather than simple copy-and-paste, is the product of the author adapting the precursory text to fit her own developing text. Whereas it might make sense to make use of short popular phrases to create amusing chapter titles, nobody writing a 165,000-word novel is going to cherry pick short phrases. As we saw in the above analysis demonstrating how the opening of Freeman's 2011 Draft was adapted to create the revised opening of her 2013 Draft, examples of mosaic plagiarism will co-occur with other examples, usually copy-and-paste. Dr. Chaski has failed to provide the initial but crucial supportive evidence of co-location.

57. Nor has Dr. Chaski made any attempt to show that any of the allegedly cited mosaic phrases are unique, let alone rare. The sharing of the phrase "the ground to open up and swallow me" (Chaski Rep. ¶ 81) in differing longer sequences is superficially the most persuasive of Dr. Chaski's ten examples, but the fact that Google finds 26,500 instances for that phrase suggests that the expression long ago became an idiom. The simple co-occurrence of such a common phrase in two texts has no evidentiary value.

58. In summary, Dr. Chaski does not acknowledge that finding pairs of phrases where some of the words are the same is only a very first step. The analyst then must demonstrate that these are actually the product of adapting precursory text, and if so that it was precursory text from the claimed plagiarised work. Dr. Chaski's small number of out-of-context examples simply does not begin to do that.

59. Finally, as noted above, the best use of identified co-occurring mosaic precursory text is to supply additional supportive evidence to help confirm a hypothesis of plagiarism first raised by the identification of unique 10+ word sequences. Without textual evidence supporting a plagiarism hypothesis in the first instance, isolated examples of alleged mosaic plagiarism are meaningless.

### C.     Conceptual Plagiarism

60. Dr. Chaski's third strategy is to claim to show "conceptual" plagiarism through a "key concept comparison." (Chaski Rep. ¶ 99.) This is initially surprising because detection of plagiarism by linguists focuses on shared language, not shared ideas—we simply do not have tools for the latter. It is therefore not surprising that Dr. Chaski: (a) does not define "concept," nor "key concept"; (b) does not explain how she discovered what the key

concepts are in the Freeman and Wolff works; and (c) in fact chose not to concentrate on "key concepts" at all. Instead, with no explanation or justification, she chose to focus instead on key**words**, which are not at all the same as key **concepts**.

61. At its simplest, a keyword is a word which is used particularly frequently in the text under consideration when **compared** with the word's general use. For instance, while "Tarot" may occur at most once or twice every million words in general usage, in *BMR* it occurs some 185 times per million. Thus, a keyword analysis can give an insight into the basic concerns of a text, and a genuine, reliable, keyword analysis could even potentially reveal valuable insights about the content and therefore the degree of shared content in a set of books. Of course, this would not in itself (especially in 165,000-word novels) demonstrate copying or plagiarism, since two works being about similar content is not in itself evidence of plagiarism. But a genuine keyword analysis could be an entry point to searching for evidence, although the conclusive evidence will always be long shared sequences that do not occur in other works.

62. However, Dr. Chaski's keyword analysis is neither genuine nor reliable because the keywords she selected were not objectively derived from the Freeman and Wolff corpora. Instead of objectively extracting keywords by computer program, Dr. Chaski admits she "selected keywords based on the complaint's description of **similarities** between the Freeman Manuscripts and the Wolff books." (Chaski Rep. ¶ 103(e) (emphasis mine).) Thus, all **differences** in keyword use were excluded from consideration. In taking keywords from arguments the Plaintiff and her lawyers had constructed for the purpose of alleging similarities, all keywords that would potentially support the contention that the Wolff books were independent creations were automatically excluded from consideration from the very beginning.

63. Beyond reliance on the Plaintiff's Complaint, there is also no explanation of **how** Dr. Chaski actually derived her 35 alleged keywords. Dr. Chaski states that she "had 35 words for keywords" based on her review of the Complaint, but how they were isolated or using what criteria is not revealed. (See Chaski Rep. ¶ 103(f).) Still, it is evident that the process was not through an independent analysis of the books.

64. Let us unpack these issues further. Firstly, the keyword data is in fact a subjective selection made by Dr. Chaski from a pre-existing subjective selection by the Plaintiff (or her lawyers)

that was aimed at identifying similarities between the two sets of texts. Among many other issues, by using the similarity comparisons in the Plaintiff's Complaint as a starting point, Dr. Chaski fails to consider words that could be keywords in the Freeman books but not in *Crave*, and, more importantly, vice versa, thereby systematically ignoring differences that would support a claim of independent authorship.

65. Secondly, Dr. Chaski does not tell us what criteria she used to select her keywords from "the complaint's description of similarities." (Chaski Rep. ¶ 103(e-f).) Some of the choices are at best surprising. Let us look at two keywords selected from the beginning of Dr. Chaski's list: "alien and "buck." (Chaski Rep. ¶ 122 tbl. 5 (Chaski's full list of the 35 chosen "keywords").) "Alien" does not occur **at all** in Freeman's 2011 Draft and only **once** in her 2013 Draft, while "buck" occurs only **once** in each—odd candidates for keywords and even odder candidates for key "concepts." Indeed, this leads one to immediately question how such words can possibly be considered "key," when the supposed keyword occurs only once or not at all in the Freeman works.

66. Meanwhile, the words "Tarot" and "priestess" **are** likely to be keywords in Freeman's works given that Tarot cards is a significant theme and that the High Priestess is often invoked. In fact, "Tarot" appears 27 times in the 2011 Draft and 29 times in the 2013 Draft, while "priestess(es)" appears 11 and 39 times respectively. Furthermore, both Freeman drafts are also very concerned with people having "visions." The following sentence from early in the 2011 Draft is illustrative: "With all of my so-called talents, knowing things, dreams and the **visions** from the **Tarot**, you'd think I would have had some inkling of my dad and grandparents' deaths." (2011 Draft p. 11.) And as it turns out, "vision(s)," appears 39 times in the 2011 Draft and 56 times in the 2013 Draft.

67. Yet these three words—"Tarot," "priestess(es)" and "vision(s)"—are absent from Dr. Chaski's list of supposed "keywords," while words that Freeman scarcely used (like "alien" and "buck") are included. Of course, "Tarot" and "priestess(es)" do not occur at all in *Crave*, and "vision(s)" only once, which perhaps provides a clue to the thinking behind the Chaski choices. Clearly, had Dr. Chaski chosen to consider them, these three items would have reinforced the evidence of significant differences between the Freeman and Wolff works.

68. A simple table which combines two of Dr. Chaski's chosen keywords, "alien" and "buck," with alternatives that are arguably more objective, will crystalise this point:

23

| Word | | 2011 Draft (*BMR*) | 2013 Draft (*Masqued*) | *Crave* |
|---|---|:---:|:---:|:---:|
| **Tarot** | [not used by Chaski] | 27 | 29 | 0 |
| **Priestess** | [not used by Chaski] | 11 | 39 | 0 |
| **Vision** | [not used by Chaski] | 39 | 56 | 1 |
| **Alien** | | 1 | 1 | 12 |
| **Buck** | | 0 | 1 | 4 |

*Table 1: Comparative Word Frequencies for Select Chaski Keywords and Potential Alternatives*[17]

69. Table 2, which adds several additional "keywords" that Dr. Chaski did use, makes a different but related point: that the Freeman/Wolff novels are actually concerned with different topics. Indeed, this becomes evident when one considers seven of Dr. Chaski's own keywords:

| Word | | 2011 Draft (*BMR*) | 2013 Draft (*Masqued*) | *Crave* |
|---|---|:---:|:---:|:---:|
| **Tarot** | [not used by Chaski] | 27 | 29 | 0 |
| **Priestess** | [not used by Chaski] | 11 | 39 | 0 |
| **Vision** | [not used by Chaski] | 39 | 56 | 1 |
| **Alien** | | 1 | 1 | 12 |
| **Buck** | | 0 | 1 | 4 |
| **(San) Diego** | | 1 | 0 | 30 |
| **Dragon** | | 1 | 3 | 104 |
| **Moon** | | 113 | 104 | 12 |
| **Wolf** | | 149 | 73 | 25 |
| **Vampire** | | 55 | 16 | 157 |

*Table 2: Comparative Word Frequencies with Additional Chaski Keywords*

70. While simple, the above table provides a useful snapshot of the differing concerns of the two authors. It is clear that *BMR* and *Masqued* have no interest in aliens or dragons, but a significant interest in Tarot, priestesses, visions, the moon, and wolves. *Crave*, by contrast, has a significant interest in dragons, no interest in Tarot, priestesses, or visions, and some

---

[17] This and the ensuing table contain only raw figures (using both singular and plural of the selected words, e.g., "Priestesses" and "wolves"). I have not attempted to correct for the fact that Freeman's 2013 Draft is a considerably shorter text than the 2011 Draft and *Crave*. Any correction to turn the 2013 Draft figures into "occurrences per 1000 words" would likely increase the figures by some 20%.

in the moon and wolves, although much less than *BMR* and *Masqued* have. Also, despite the Complaint highlighting the fact that "San Diego" is important in both sets of texts, *Crave* mentions the city fairly often, while *BMR* mentions it only once and *Masqued* not at all. Finally, all three have an interest in vampires, but, while "vampire" is easily the most prominent keyword in this particular list for *Crave*, it has much less significance in Freeman's 2011 and 2013 Drafts than do "Tarot," "priestess," "vision," "moon," and "wolf."

71. In summary, several points are now clear:

    a. Firstly, Dr. Chaski's "keyword" list is not in fact a list of keywords as standardly understood; neither are they keywords of the individual books nor of the composite set. The fact that Dr. Chaski did not use "Tarot," "priestess," or "vision," even though they appear quite frequently in *BMR*/*Masqued*, confirms that her selections were not based on objective sampling. While a genuine, reliable keyword analysis potentially allows insight into the main concerns of the books, there can be no value in any results derived from the analysis of data derived from lexical clusters attached to Dr. Chaski's subjectively selected, pseudo "keywords."

    b. Secondly, even when one is focusing on Dr. Chaski's selected "keywords," the actual word frequencies themselves indicate differences between the texts, not similarities.

72. Finally, the foregoing illustrates what an objective analysis of keyword frequencies can potentially contribute to text comparison—it can provide some data as to the subject matter of different works (although this would be just another data point to consider in a plagiarism analysis, and not conclusive). Note, however, that this is not how Dr. Chaski used her keywords. Instead of examining word frequencies as I have done above, Dr. Chaski "created lexical clusters for each of these keywords using only lemmatized content" and then compared such lexical clusters with those from a set of alleged baseline novels. (Chaski Rep. ¶ 103(g)-(n)).) The value of this exercise is unclear. It is not an established procedure—in essence its employment here is an experiment. As already noted above in my ¶ 7, the way

Dr. Johnson (1997) used keywords is very different and provides no academic support to Dr. Chaski.[18]

73. But even were the keywords correctly isolated, Dr. Chaski's analysis of the "overlap rate" of "lexical clusters" would have been irremediably compromised by the nature of the corpus of "baseline" documents that she used in her analysis. While the use of a comparison corpus of texts is standard practice in computational linguistics, it is usual for the linguist herself to choose the items that make up the corpus and to justify her choices. Indeed, selecting and justifying her baseline choices is something Dr. Chaski undertook convincingly in the *Yukos* case, in which we were both involved (see Section II.D above for details). In this case, however, the set of baseline documents was selected not by Dr. Chaski, but apparently by the Plaintiff and/or her lawyers. (See Chaski Rep. ¶ 42, with table listing "the documents which I **received** for examination in this case," which includes the 10 baseline novels.) Worryingly, Dr. Chaski provides no indication of the criteria by which these comparator books were selected, nor whether she was satisfied that the selection was objective and unbiased and the corpus sufficiently large.

74. Therefore, let us look more closely at the corpus. As *Crave* is recognised as a book about vampires and set in a boarding school in Alaska, there are some immediately noticeable oddities about the corpus. According to the keyword cluster data provided in Dr. Chaski's report (Chaski Rep. Annex 3, pp. 119-154) it would appear that only six of the ten chosen books are actually about vampires, which are central to *Crave*. On the other hand, it appears that five are about aliens, which are not mentioned at all in *BMR*, only once in *Masqued*, and rarely in *Crave*. It also appears that none of the comparator books is set in Alaska, and that only two of them even mention Alaska, even though "Alaska" is one of Dr. Chaski's chosen keywords. Certainly, there is no attempt to argue that these ten books are an objective or robust sample, notwithstanding the claim by Dr. Juola (who apparently used,

---

[18] In footnote no. 1 of her report, Dr. Chaski states that Johnson (1997) "presents a method of lexical overlap rates on keywords" and that "[t]he only difference between Johnson's method and the method used in this report has to do with the additional statistical analysis I present." (Chaski Rep. ¶ 2 n. 1.) Having reviewed Dr. Johnson's paper cited by Chaski (and being very familiar with Johnson's work), there are substantial differences in the basic methodology. Most notably, Dr. Johnson's paper contains no equivalent use of any form of lexical clustering; what is being measured is phrasal overlap, whether complete or partial. Furthermore, Johnson's control set is taken from three other students answering the same question, not a selection of authors writing in the same or a claimed similar genre. Johnson is also the sole selector of the data to be examined and of the control set, whereas Chaski depends on third parties for both the data from the Complaint and for the control set from unknown sources. Taken together, these differences demonstrate that Dr. Chaski's assertion that she is using a similar method to Johnson is simply a misrepresentation.

but also did not select, the same set) that "[t]hese additional novels provided an objective baseline to let us control for the specific characteristics of this genre." (Juola Rep. ¶ 15.)

75. Finally, I will not discuss the apparently sophisticated statistical tests that Dr. Chaski applied to her results (beginning Chaski Rep. ¶ 108) because the input data were flawed and consequently the results are worthless.

### D.   Concluding Evaluation of Dr. Chaski's Report

76. All three components of Dr. Chaski's analysis are unreliable. Firstly, for "copy-and-paste" plagiarism, Dr. Chaski mistakenly concentrated on 6-word sequences, which are too short to count as reliable evidence of copying. Then, she failed to try to confirm, by a Google search, that the sequences she considered were unique to the two sets of books. Because of this, Dr. Chaski lacks the data needed to even float a plagiarism hypothesis.

77. Secondly, for "mosaic" plagiarism, Dr. Chaski initially demonstrated only that Wolff produces intertextual chapter titles like "Carpe Kill-Em." That is not evidence that she was also adapting precursory Freeman text in order to incorporate it into her books. Dr. Chaski supported her claim about "mosaic" plagiarism with only ten examples, which were taken out of context, apparently borrowed directly from the Plaintiff's Complaint, and not linked to show an attempt to conceal borrowing.

78. Finally, Dr Chaski's "key concept"/keyword analysis has no validity. Even if such an untried approach could prove insightful in some situations, Dr. Chaski's analysis (and her dependent statistical analysis) has no value because neither her "keywords" nor her "baseline" works were objectively selected.

## V.   Evaluation of Dr. Juola's Expert Report/Opinions

79. Dr. Juola had great success, some 10 years ago, using computational linguistic analysis to identify J.K. Rowling as the author of a book published under the name of Robert Galbraith. For this reason, the reader would assume that, in this Freeman authorship case, Dr. Juola would have adopted a similar strategy: i.e., using his famous Java Graphical Authorship Attribution Program (JGAAP) (see Juola Rep. ¶ 6) to first characterise the Freeman style and then use this analysis to test whether the scores for such features as word length, word frequency, patterns of punctuation, etc. identified Freeman's linguistic fingerprint in

(sections of) the *Crave* texts.[19] But he chose not to employ that program or methodology here.

80. In fact, the report Dr. Juola has submitted is something of an aberration. Not only did Dr. Juola not use his JGAAP program (as he did in investigating Rowling/Galbraith), he did not even collect all the data to be analysed himself. Instead, he relied on some pre-analysed data produced by Mr. Trent Baer, the Plaintiff's husband. Dr. Juola states: "[I] took advantage of Baer's prior work in analyzing the relevant books." (Juola Rep. ¶ 21.) Much worse, like Dr. Chaski, Dr. Juola bases much of his analysis on a misreading of publications by myself and my colleague Dr. Johnson.

81. On the positive side, one very important decision that Dr. Juola made, but that sadly was not made by Dr. Chaski, was to avoid oversampling. He states: "For Blue Moon Rising, to avoid oversampling non-independent texts (different versions of the same work), we confined our analysis to the 2011 manuscript." (Juola Rep. ¶ 15.) One problem with oversampling is that the "same" item can be counted multiple times if it re-occurs in various versions of the same basic text. As we do not have her raw data, we do not know whether Dr. Chaski's claimed 700 6-word items included repeats, thereby falsely exaggerating their evidentiary significance. However, although Dr. Juola was initially correct to avoid oversampling, he departed from reliable principles of Forensic Linguistics in the rest of his report.

### A.      7-Word Sequences

82. Firstly, like Dr. Chaski, much of Dr. Juola's analysis is based, crucially and fatally, on the mistaken premise that shared 6- and 7-word sequences are in themselves solid evidence of plagiarism. More worryingly, like Dr. Chaski, Dr. Juola claims to derive this faulty assumption from my own work.

83. Specifically, Dr. Juola writes: "Coulthard has noted that 'the occurrence of long identical sequences in two texts is likely to be a product of borrowing' [footnote to Coulthard, et al., 2017, p. 189], and that a sequence of only seven English words is unlikely enough to be a unique phrase…." (Juola Rep. ¶ 14 & n. 8 (citing the 2017 edition of my book with Dr.

---

[19]     See   https://www.scientificamerican.com/article/how-a-computer-program-helped-show-jk-rowling-write-a-cuckoos-calling/ (Dr. Juola discussing his methodology and use of his JGAAP program in investigating Rowling/Galbraith).

Johnson and David Wright, *Forensic Linguistics: Language in Evidence*).)[20] However, that is not what our 2017 book actually said. We did note that "the occurrence of long identical sequences in two texts is likely to be a product of borrowing," but we were referring to significantly longer sequences. More important, what Dr. Juola says subsequently regarding the evidentiary value of "only seven English words" is not quoted from and certainly not endorsed by me, and is in fact dangerously wrong. Had Dr. Juola continued reading to the end of the section of our 2017 book from which he quoted—incidentally headed "*The evidential value of single identical strings*"—he would have read this conclusion: "From evidence like this we can assert that even a sequence **as short as 12 running words** [can] have a very high chance of being unique occurrence…."[21] Thus, Dr. Juola's reliance on 7-word sequences as evidence of plagiarism has no basis in the research literature from which he claims to draw support.

84. But even if Dr. Juola's 7-word starting point were legitimate, the results he presents are inadequate as evidence. He begins his analysis by saying: "We first inspected the text of … both Crave and BMR to see if there were any word-for-word overlaps of seven words or more that did not admit of an obvious explanation." (Juola Rep. ¶ 17.) Dr Juola does not say how he conducted this inspection, whether by eye/hand or computer program, but what is most notable is the paucity of the evidence he adduces. Unlike Dr. Chaski, who claimed hundreds of examples (but did not produce them), Dr. Juola cites only **three** examples that he deems worthy of consideration. He says: "We did find several such phrases shared between Crave and BMR, of which 'in the air as I try to', 'that million-dollar smile of his', and 'on my arms and the back of my neck' are the most notable." (Juola Rep. ¶ 19.) To base an accusation of plagiarism on three isolated phrases is to say the least unusual, but let us examine these three "notable" examples in turn:

   a. **"that million-dollar smile of his"**: This is actually only a 6-word sequence (as Dr. Juola acknowledges later in ¶ 19), and could be considered even shorter because it includes a hyphen indicating that "million" and "dollar" are not separate choices, but rather selected as one item—indeed the idiom "million-dollar smile" could even

---

[20] M. Coulthard, A. Johnson, and D. Wright, *Forensic Linguistics: Language in Evidence*, 2nd edition. London: Routledge (2017).
[21] *Id.* p. 190 (emphasis mine).

been seen as a single choice. Then, Dr. Juola acknowledges that "that million-dollar smile of his" is in fact quite a common expression, appearing "3,370 times" on Google (Juola Rep. ¶ 19), and as such not reliable evidence for plagiarism.

b. **"on my arms and the back of my neck"**: Dr. Juola indicates that he believes this second phrase is stronger evidence, noting that it appears in "roughly 20 independent contexts in Google search, which is rare." (Juola Rep. ¶ 19.) However, a Google search for that exact phrase as of August 23, 2023 in fact yields some 320,000 results, making it not rare at all. The expression is also made up of two common idioms, "the hair on my arms" and "the back of my neck." Again, not reliable evidence for plagiarism.

c. **"in the air as I try to"**: Dr. Juola moves on to this third short phrase, which he claims "appears only seven times in this set." (Juola Rep. ¶ 19.) It is unclear what "set" he is referring to because the phrase in question actually appears roughly 1,400 times in Google Books (which Juola sometimes uses as a reference) and roughly 500,000 times in ordinary Google (all as of August 23, 2023). Dr. Juola concludes his discussion of this phrase by saying: "Thus, 'in the air as I try to' qualifies as a seven-word lexical overlap and is, in Coulthard's view, unlikely to be independently written by two different authors." (Juola Rep. ¶ 20.) In which case Coulthard would have been very wrong, if he had indeed relied on 7- and not 10-word sequences, given that there are thousands of occurrences of the phrase in Google Books and hundreds of thousands in an ordinary Google search. Again, not reliable evidence for plagiarism.

85. To summarise so far, Dr. Juola has provided three phrases that occur in both *BMR* and the Crave series, 9, 7, and 6 words long. But a cursory consultation of Google shows none of the three is even close to rare, let alone unique to the two authors. This evidence can be disregarded.

## B. Even Shorter Sequences

86. Having cited as evidence only two phrases that were at least 7 words long—which was his original, though mistaken, entry requirement—Dr. Juola moves on to focus on even shorter phrases, which were apparently presented to him in a document authored by Mr. Baer, to which I do not have access. (Juola Rep. ¶ 22 (mentioning "six commonalities related to the

30

2011 version identified by Baer").) Dr. Juola then focuses, in ¶ 22, on four of the Baer "commonalities," one of which he has already discussed ("million-dollar smile"):

    a. **"tree stump seats"/"tree-stump seats"**: He notes that the phrase "tree stump seats" appears in *BMR* and "a similar phrase appears as 'tree-stump seats' in the Crave series," though the two are not identical;

    b. **"small Stonehenge"/"Stonehenge lite"**: He notes that the words "small Stonehenge" appear in *BMR* and the words "Stonehenge lite" appear in *Crave*. But again, the two phrases are not identical and this time the books share only a single word;

    c. **"air as I try to"**: He notes that the phrase "air as I try to" appears in both, although he admits that they occur "in wildly different contexts," which counts as an admission that plagiarism is unlikely;[22]

    d. **"that million-dollar smile of his"**: He notes that this phrase appears in both, but also says "note hyphen," with no further comment.

87. There is a sense of desperation here. Having first (wrongly) claimed that 7-word sequences were generally agreed to be evidence of precursory text, Dr. Juola now uses as evidence: (a) one 6- (or 5-) word example ("that million-dollar smile of his"), (b) one 5-word example ("air as I try to"); (c) one 3-word example ("tree stump seats"), although this could be challenged as a match given that the hyphen is missing in the *Crave* example and Dr. Juola thought the hyphen sufficiently important to note its presence in "million-dollar"; plus (d) a **1-word** example **(**"Stonehenge"), which occurs in two different linguistic contexts.

88. In order to claim that these four short phrases have any evidentiary value, Dr. Juola sets out to demonstrate that they are "uncommon phrases," and does so by searching for each of them in Google Books N-grams (abbreviated "GBN"). (Juola Rep. ¶ 25.) He notes that the four examples were "not occurring in the Google Books Ngrams viewers," which, he argues, makes each "an uncommon phrase." (Juola Rep. ¶¶ 22, 25.) Dr. Juola's underlying argument is that because these four short phrases (a) appear in *BMR*; (b) also appear in

---

[22] Specifically, Dr. Juola shows that this phrase is used very differently as follows: "my arms ailing wildly in the air as I try to catch my balance" (*BMR*); "I'm gasping for air as I try to make it through the roiling clouds" (*Crave*); "The words hang in the air as I try to absorb them" (*Crave*). (Juola Rep. ¶ 22 (third bullet).) I agree that the phrase is used in "wildly different contexts," though unlike Dr. Juola I appreciate that this makes borrowing that phrase even less likely.

*Crave*; but (c) do not appear in Google N-grams, he can reliably rule out the possibility of independent creation. He even attempts to put a probability into the finding: "Continuing our analysis, we calculate the probability of finding a word in BMR that is also in Crave, but that is not in GBN to be a tiny fraction of a percent, substantially less than one chance in 1000." (Juola Rep. ¶ 28.)

89. However, I know of no published scholarship that justifies and validates the use of the Google N-grams database in plagiarism detection, and Dr. Juola himself does not provide references to any such work, other than another spurious reference to myself. He does this by taking a genuine quote from myself (discussed in my ¶ 83 above) and adding his own assertions, the latter of which I have highlighted in bold below:

> Coulthard has noted that "the occurrence of long identical sequences in two texts is likely to be a product of borrowing" [footnote to my 2017 book], and that a sequence of only seven English words is unlikely enough to be a unique phrase, **even against the entire contents of the Google search engine database or Google Books Ngram database.**

(Juola Rep. ¶ 14 (bold emphasis mine to show where Dr. Juola has inaccurately attributed his proposition regarding Google N-grams to me).)

90. We can disregard any evidentiary claims based on Google N-grams, as this is not a recognised resource or measure of the uniqueness of short phrases. I note in passing that Google N-grams is intended as a tool to "explore language usage trends over time,"[23] not to detect plagiarism, and also that, as Dr. Juola later acknowledges himself in his technical appendix, "A word or phrase only appears in GBN if it has at least forty occurrences in the 15,000,000 books (roughly 150,000,000,000 words) database of Google Books." (Juola Rep. ¶ 43(a).) So, "not occurring in the Google Books Ngrams viewer" does not mean what Dr. Juola claims.

91. To illustrate this: Dr. Juola has implied that "small Stonehenge" is a rare phrase because he did not find it in Google N-grams. However, "small Stonehenge" appears over 300 times in Google Books and over 8,000 times in ordinary Google.[24] So, it is not a rare phrase at all. Interestingly, when one searches Google Books for Wolff's phrase "Stonehenge lite," the

---

[23]https://guides.library.upenn.edu/penntdm/tools/ngram_viewer#:~:text=Google%20Ngram%20Viewer%20is%20a,trend%20of%20terms%20over%20time.
[24] All results here as of August 21, 2023.

**only** result is her own book, *Court*. In other words, Freeman used the relatively **common** phrase "Small Stonehenge," while Wolff used the rare "Stonehenge lite," which Google Books attributes only to her.

92. Similarly, as to the other phrases which Dr. Juola said were no-shows in Google N-Grams:

    a.  "air as I try to":          Google Books 2,190 and Google 588,000;

    b.  "that million-dollar smile of his":  Google Books 182 and Google 4,170;

    c.  "tree stump seats":          Google Books 225 and Google 13,900.

93. In summary, Dr. Juola used an inherently limited and therefore inappropriate database to investigate whether the phrases of interest were uncommon. Had he used the more robust Google Books database, or the even more comprehensive ordinary Google, he would have seen that none of the four phrases is rare, and that several are in fact quite common. Since Dr. Juola cannot reliably conclude that the four phrases at issue are rare as he alleges, his entire analysis (including his statistical analysis) falls apart.

94. As with Dr Chaski's results, there is no need to evaluate Dr. Juola's statistical analysis, because it is similarly premised on very limited and collectively meaningless data for all of the reasons just explained.

95. Dr. Juola then argues for the evidentiary value of a sequence of **four letters**, "katm." He notes: "My attention has been drawn to the proper names "Katmere" (Crave) and "Katmai" (Freeman copyrighted The World.doc note) which both contains (sic) the highly unusual character cluster 'katm.'" (Juola Rep. ¶ 34.) Putting to one side the oddity of claiming plagiarism for an item that does not occur in any versions of the supposedly plagiarised books (apparently it was found only in a set of Freeman's notes), as a linguist Dr. Juola should have immediately drawn attention to the absurdity of arguing that mutual use of a single four-letter cluster is evidence of plagiarism. Dr. Juola also failed to consider that "Katmai" is a real-world Alaskan park,[25] which perhaps explains its use by Freeman. On the other hand, I have been reliably informed that "Katmere Academy" was chosen as an intertextual reference to the school in the Harry Potter books, Hogwarts. We know from Dr. Chaski's evidence that Wolff likes playing with intertextual links, and we discover from Wikipedia that "theory has it that Rowena Ravenclaw came up with the name of Hogwarts

---

[25] https://www.nps.gov/katm/index.htm.

after dreaming of a warty hog that led her to a cliff by a lake."[26] Similarly, the name of the Crave boarding school can be seen as a reference, indeed homage, to J.K. Rowling's reversal of 'warthogs' to become "'Hogwarts."  Here, the (near) reversal of 'meerkat' has produced the Katmere Academy.[27]

96. Dr. Juola's final contribution is that "we collected ten different YA vampire novels by ten different authors and confirmed that neither 'tree stump seats,' 'Stonehenge lite,' 'air as I try to,' nor the 'katm' letter cluster appear anywhere in these ten novels." (Juola Rep. ¶ 36.) Based on this, Dr. Juola "concludes that these phrases are not an integral part of the supernatural fiction genre." (*Id.*) It is difficult to know where to begin. Firstly, why would anyone assume these four items would be an integral part of the genre, and why would that matter? Secondly, Dr. Juola's own report shows that the "katm letter cluster" and "Stonehenge lite" do not occur in *BMR*/*Masqued*. Thirdly, "tree stump seats" does not occur in *Crave* at all and only in the other books in the series in its hyphenated form. Fourthly, he does not give details about how "we collected ten different YA vampire novels." However, these are the same ten novels that Dr. Chaski used, and she admits that they were actually provided by the Plaintiff or her lawyers. (Chaski Rep. ¶ 42 (listing the 10 books among "the documents which I received for examination in this case").) Like Dr. Chaski, it would appear that Dr. Juola did not actually select them using objective criteria or specifically for his own analysis.

97. Dr. Juola concludes his report by conceding that it is "not unbelievable" that Wolff authored the Crave series independently and by accepting that the alleged similarities could have come from external sources. (Juola Rep. ¶ 37 ("while it is not unbelievable that the Crave series and BMR are independently written, it is more likely than not that they both derive from the same specific sources").) We can go much further: Dr. Juola has produced no evidence that reliably indicates that Wolff copied even one word from Freeman.

---

[26]https://en.wikipedia.org/wiki/Hogwarts#:~:text=Theory%20has%20it%20that%20Rowena,other%20wizarding%20schools%20and%20Muggles.
[27] I was directed to a published Q&A where Wolff explained: "Katmere is actually my homage to Harry Potter and the Lion King combined. Hogwarts is Warthog rearranged, so Katmere is Meerkat rearranged." https://www.google.com/books/edition/Katmere_Academy_An_Insider_s_Guide/QKxIEAAAQBAJ?hl=en&gbpv=1&dq=%22katmere+academy%22+%22meerkat%22&pg=PT138&printsec=frontcover.

C.          **Concluding Evaluation of Dr. Juola's Report**

98.  Dr. Juola mistakenly relied on 7-word sequences and used the Google N-grams viewer as
an investigative tool, but did not cite any published research which genuinely supported
these decisions—although he did misquote my own work as pseudo-support. Given the
erroneous basis of Dr. Juola's initial 7-word analysis, the paucity of the data he manages to
accumulate, and his exceedingly weak conclusions, Dr. Juola has comprehensively failed
to produce any evidence of plagiarism and his report should be disregarded.


I declare under penalty of perjury that the foregoing is true and correct to the best of my knowledge
and belief.



_____

Dr. Malcolm Coulthard

# APPENDIX A

# Dr. Malcolm Coulthard
## A Short Forensic *Curriculum Vitae*

### PROFESSORSHIPS
- Emeritus Professor of Forensic Linguistics, Aston University
- Emeritus Professor of English Language and Linguistics, University of Birmingham
- Honorary Professor, School of English, University of Cardiff

### DEGREES
- BA in English, with First Class Honours, University of Sheffield, 1964
- Postgraduate Certificate of Education, London Institute of Education, 1965
- Diploma in Phonetics, University College, London, 1966
- MA in General Linguistics, University College London, 1967
- PhD in English Language and Linguistics, University of Birmingham, 1970

### POSITIONS
- Research Fellow in Arts, University of Birmingham, 1967-1970
- Lecturer in English Language, University of Birmingham, 1970-1984
- Senior Lecturer in English Language, University of Birmingham, 1984-1993
- Professor of English Language and Linguistics, University of Birmingham, 1993-2004
- Professor of Forensic Linguistics, Aston University, 2004-2011
- Visiting Professor, Federal University of Santa Catarina, Brazil, 2012-2017

### RESEARCH
- Since 1970, I have been continuously involved in research into the analysis of verbal interaction and of written text. I have conducted individual and team research and published 34 books and special editions of journals and some 100 articles and book chapters. According to Google Scholar there are 23,758 citations of my publications and I have an 'h-index' of 42 and an i10 index of 24 ('an h-index of 40 is outstanding' and 'an i-10 index of 25… is … excellent'). I have supervised and examined some 120 MA and PhD theses in the areas of text and discourse analysis, translation, and forensic linguistics.

- Twenty of my ex-doctoral students have held university posts in Brazil, Chile, England, Hong Kong, Portugal, and Wales, four of whom reached the rank of full Professor and four more became Associate Professors/Readers.

### ACADEMIC HIGHLIGHTS
- My last position before retirement was at the Federal University of Santa Catarina, where I was employed primarily to teach the analysis of written and spoken text and forensic language analysis. I established Forensic Linguistics as an academic subject in Brazil and was elected Honorary President of the newly formed Association of Language and Law, (ALIDI).

- I am Emeritus Professor at the University of Birmingham, where in the 1990s, I developed Forensic Linguistics as an undergraduate subject.

- In 2001 I was appointed Honorary Professor at the University of Cardiff, where I helped to set up the world's first Master's degree in Forensic Linguistics, which is still in good health.

- I am also Emeritus Professor at the University of Aston, where in 2008, I established the world's first Centre for Forensic Linguistics with two members of staff. It now teaches forensic linguistics at Undergraduate and Postgraduate levels and hosted the Biennial conference of the International Association for Forensic and Legal Linguistics twice, in 2011 and in 2021. Now renamed as the Aston Institute for Forensic Linguistics, it has a permanent staff of eight and has recently won a £6.7 million, ($8 million), five-year government grant to research five separate areas of forensic linguistics and employ 11 post-doc researchers.

- I have travelled widely to teach courses, to lecture, to present research papers and to examine in: Argentina, Australia, Belgium, Brazil, Chile, Croatia, Cyprus, Czech Republic, Denmark, Egypt, Finland, France, Germany, Greece, Holland, Hong Kong, Iran, Ireland, Israel, Italy, Malta, Japan, Malaysia, Poland, Portugal, Singapore, Spain, Sweden, Tunisia, Uruguay, the USA, and Venezuela.

- In 1993 I was the Chair of the Founding Committee of the International Association of Forensic Linguists, (now renamed the International Association for Forensic and Legal Linguistics) and became its first President.

- I was also the founding co-editor of the two main forensic linguistics journals: *International Journal of Forensic Linguistics* (1994) and *Language and Law - Linguagem e Direito* (2015).

- I organised the First British Seminar on Forensic Linguistics in March 1992, the Second British Seminar on Forensic Linguistics in 1993 and the Fourth Biennial Conference of the International Association of Forensic Linguists in 1999, all of them hosted by the Centre for Advanced Research in English at Birmingham University, of which I was the founding Director. Since I went to Aston the group has hosted the Conference in 2011and 2021, and while I was working in Brazil, we hosted international conferences in 2013 and 2018.

- Since 2000, along with an ex-student Dr. K Kredens, I have run the annual International Summer School in Forensic Linguistic Analysis, which from 2006 to 2011, was hosted by Aston University, the most recent summer school was in Manila in July of 2023.

**FORENSIC WORK**

- I have been commissioned to prepare reports for both prosecution and defence in over 230 cases including 8 reports for the Criminal Cases Review Commission, 1 for the Scottish Criminal Cases Review Commission, 1 for the Police Ombudsman in Northern Ireland and 2 for the financial detective companies Kroll and Control Risks. I wrote reports for the following UK appeals against conviction: the Birmingham Six, Derek Bentley, Paul Blackburn, the Bridgewater Four, Robert Brown, Reg Dudley and Ian Maynard, Iain Hay Gordon, Paul Malone, Peter Tredget, and the Campbell 'Glasgow Ice Cream Wars' Appeal.

- I have given evidence in court many times, including a case of academic plagiarism in the High Court in Hong Kong, three terrorist trials in Northern Ireland, a murder trial in Norwich, the Danielle Jones/Stuart Campbell 'texting' murder trial in Chelmsford, a Court Martial in Germany and the Derek Bentley, Paul Malone, Robert Burton, Paul Blackburn, and Peter Tredget Appeals in the Royal Courts of Justice in London. I have also worked with the Metropolitan, the Scottish, the South Wales, and the Military police on internal investigations. My most recent court appearance was at in the Peter Tredget Appeal in October 2021. My most recent completed authorship cases were the $50 billion (sic) Yukos Arbitration case, which was settled in the Hague in February 2020 and a successful Appeal in Brazil for which I produced evidence of plagiarism in the final Judgment.

# APPENDIX B

# Emeritus Professor Malcolm Coulthard
# Publications List (2013-2023)

## FESTSCHRIFTS

- *Linguagem & direito: caminhos para a linguística: Uma homenagem a Malcolm Coulthard.* Virgínia Colares (ed) São Paulo, Cortez, 2016

- A Linguagem E A Lei: Trajetórias Em Homenagem ao Professor e Pesquisador Dr. Malcolm Coulthard, M J Martins and C P Hagemeyer (eds) Ananindeua, Editora Itacaiúnas 2023

## THESIS AND AUTHORED BOOKS

- An *Introduction to Forensic Linguistics: Language in Evidence*, 2nd ed. (with Johnson, A and Wright, D), London, Routledge 2017.

## EDITED BOOKS AND SPECIAL ISSUES OF JOURNALS

- *Analise do Discurso*, (with Caldas-Coulthard C and Heberle V) Florianopolis, DLLE/CCE/UFSC 2013

- *Studies in Discourse Analysis (with M* Montgomery) *(2nd ed.)* London, Routledge 2013.

- *Echoes*: *Reflections on Language and Literature* (with Tumolo C and Beck M) Florianopolis. PGI/UFSC, 2014,

- *International Journal of Law, Language and Discourse*, (Guest Editor with Cheng W), 2014 Vol 3(2).

- *Linguagem e Direito: os Eixos Tematicos*, (with Colares V and Sousa-Silva R) *2015*

- *The Routledge Handbook of Forensic Linguistics*, 2nd ed (with May, A and Sousa-Silva R), London, Routledge 2020

- *Perspectivas em linguistica forense*, (with de Almeida, D, and Sousa-Silva, R) Campinas, SP, Unicamp/Publiaçoes IEF, 2020

- *Methodologies and Challenges in Forensic Linguistic Casework,* (with Picornell, I, and Perkins, R, Chichester, Wiley Blackwell 2022

- *Texts and Practices Revisited* (with Caldas-Coulthard, C R), London, Routledge, 2023

## JOURNAL EDITING

- Founding editor *Language and Law – Linguagem e Direito* 2013-

- Member of the editorial board of *Text* since its foundation in 1981

- Member of honorary board of *Discourse Studies*, since its foundation in 1999

- Member of the editorial board of the *Applied Linguistics*, 2002-6

- Member of the editorial board of the *Journal of Applied Linguistic and Professional Practice*, 2003-

**PUBLISHED ARTICLES AND CHAPTERS IN BOOKS**

- 'The Linguist in the Witness Box 'in Candlin C and Sarangi S (eds) *Handbook of Communication in Organisations and Professions* The Hague, Mouton, 2011591-608

- 'The Linguist in the Witness Box', (in press) in Candlin C and Sarangi S (eds) *Communication in Professions and Organizations,* The Hague, Mouton

- 'Forensic Linguistics' (in press) entry in the *Linguistics Encyclopedia*, 3rd edition, London, Routledge

- 'On the admissibility of Linguistic Evidence', *Journal of Law and Policy,* 2013 23, 2, 441-468

- 'PERIGO, CUIDADO, ATENÇÃO: a comunicação linguística de risco em advertências de produtos', (with Hagemeyer C) *Cadernos de Linguagem e Sociedade,* 2013, 14 (2) 28-53

- 'Have you been Warned?' In Casesnoves, R, Forcadell, M and Gavaldà, N (eds) '*Ens queda la paraula. Estudis de lingüística aplicada en honor a M. Teresa Turell*, Barcelona, IULA, Universitat Pompeu Fabra, 2014, 253-64

- 'Co-Writing Justice; an interview conducted by Adamantia Karali', in *This Century's Review*, 2014, issue 3, 34-41

- (2014) 'Plagiarism in the Academic Context: An investigation of PPGI students' awareness of the problem'. (with Abreu, B) In: Coulthard, Tumolo and Beck. (Orgs.). *Echoes*: *Reflections on Language and Literature*. Florianopolis: PGI/UFSC, p. 71-80.

- 'You have been Warned – or have you?' (with Hagemeyer C) in: *Crossing Boundaries: Interdisciplinarity in Language Studies* 2014.

- 'Plágio no meio acadêmico' Abreu, B em *Linguística Aplicada em Contextos Legais,* Sonia Bittencourt Silveira, Caroline Scali Abritta e Amitza Torres Vieira (Orgs.), São Paulo, Paco Editorial, 2015 357-380

- 'The Sounds of Silence' in: *Speaking of Language and Law: Conversations on the Work of Peter Tiersma,* L Solan, J Ainsworth and R Shuy, Eds. Oxford, Oxford University Press, 2015, 122-6

- 'On Product Warnings' (with Hagemeyer C) *Language and Law / Linguagem e Direito*, 2015, 2.1, Porto, Biblioteca da Universidade de Porto,

- 'Linguística Forense na Era da Tecnologia', *Língua e Literatura na Época da Tecnologia* Mota, M, Corseuil, A, Beck, M and Tumolo, C *(eds)* Florianópolis, Editora UFSC, 2015, 291-306

- 'Linguistica Forense' (with Sousa-Silva R) in *Ciências Forenses* eds Dinis-Oliveira, R and Magalhães, T, Lidel, Porto, 2016, 137-44,

- Forensic Linguistics (with Sousa-Silva R) in *What are Forensic Sciences? Concepts, Scope and Future Perspectives,* eds Dinis-Oliveira, R and Magalhães, T Lidel Group, Lisbon, 2016

- Algumas aplicações forenses da Linguística Descritiva, em *Linguagem & direito: caminhos para a linguística: Uma homenagem a Malcolm Coulthard.* Virgínia Colares (org.).  São Paulo: Cortez, 2016, pp17-48

- On the legal status of an interpreted confession (with Cal-Meyer P) in *Language and Law / Linguagem e Direito*, Vol 4.1, 1-16

- 'In my opinion', in Coulthard, M, May, A and Sousa-Silva R (eds) Routledge Handbook of Forensic Linguistics, London, Routledge 2020

- 'Detecting faked texts', (with Picornell, I) in Picornell, I, Perkins, R and Coulthard. M (eds) *Methodologies and Challenges in Forensic Linguistic Casework,* 114-28.

- 'The official version' in *Texts and Practices Revisited,* C R Caldas-Coulthard and M Coulthard (eds) 2023 99-112

# APPENDIX C

# FORENSIC LINGUISTICS IN BRAZIL: RESEARCH AND PRACTICE

**Abstract**

This chapter provides an overview of current Forensic Linguistics research and practice in Brazil, with examples drawn from casework. The intention is to show how Linguists working as Forensic Language Experts are making a growing contribution to the administration of justice in Brazil. The data analysed ranges from audio recordings of interactions in police stations through emails and contracts to court judgments. Additionally, we discuss some of the specific problems involved in legal translation illustrated with a focus on rogatory letters. The chapter draws on research carried out in Brazilian Universities and its practical application in real cases through expert forensic linguistic reports.

**Key words:** Forensic Linguistics; Language as Evidence; Legal Language; Legal Translation; Expert Reports.

# FORENSIC LINGUISTICS IN BRAZIL: RESEARCH AND PRACTICE

## 1. Introduction[1]

In Brazil, Forensic Linguistics is still in the establishing phase. Over the past 10 years, there has been a growing interest in Forensic Linguistic studies, not only within the Linguistics community, but also within the Law community, more exactly from Law students and legal professionals. Forensic Linguistics began as an academic discipline in Brazil in 2012 with a master's level course taught to some 20 students from Linguistics, Translation Studies and Law at the Federal University of Santa Catarina (UFSC). Following the course, some students registered as the first in the country to take a Doctorate in the area. Their PhD research topics covered Plagiarism, Legal Translation, the Discourse of Police Interviews in cases of violence against women and the Language of Product Warnings.

Now there is a small, but growing, number of qualified professionals who are researching in the area, teaching modules to undergraduates and beginning to act as experts (*peritos*). Since 2021, Brazilians have had the option of studying for a postgraduate qualification at a distance - the Pontifical Catholic University of Minas Gerais has an online module linked to an undergraduate Law course; the Federal University of Rio Grande do Norte (UFRN) has an online specialization degree; Unicentro, in Paraná, has offered undergraduate FL modules since 2018 and is now one of three Brazilian Universities planning postgraduate diploma courses. In addition, the University of Porto, in Portugal, offers a highly regarded distance postgraduate course from which several Brazilians have already graduated. Sadly, although in our definition Forensic Linguistics includes the areas of translation and interpreting, there is still no postgraduate degree in Brazil devoted exclusively to Legal Translation and/or Interpretation,

---

[1] We would like to thank the two anonymous reviewers who commented on an earlier version of this chapter. As a consequence, this is a much revised, shortened and, we hope, improved version. One suggestion, however, we did not adopt: 'exclude translation and interpreting' on the grounds that they do not belong within Forensic Linguistics. We resisted this suggestion because both areas are very much part of forensic linguistics practice in Brazil: Brazil is a member of the Mercosul trade association and so Brazilian international contracts are typically bilingual and contract disputes often revolve around translation problems; Brazil is a major tourist venue and Brazilian police and courtrooms need to draw regularly on the services of interpreters. But, more generally, a major section of the international community accepts these two areas as part of the discipline: the biennial conferences welcome papers and indeed plenary presentations in these areas, the main association, IAFLL, accepts professional translators and interpreters as members; the association's two journals publish articles in both areas and the *Routledge Handbook of Forensic Linguistics* includes chapters.

although currently efforts are being made to introduce them into the postgraduate course in Translation Studies at UFSC.

In 2012, a Forensic Linguistics research group, the *Grupo de Linguística Forense* (GLF*)* was founded at UFSC, consisting of established researchers and postgraduate students and the group organized two important international conferences (2013 and 2018). Ideally every new academic area has both a dedicated research publication and an academic association; the GLF was responsible for the foundation of the bilingual journal *Language and Law–Linguagem e Direito* (open access at https://ojs.letras.up.pt/index.php/LLLD), which is now hosted by the University of Porto. The journal is in its tenth year and has published over 100 articles, more than a third in Portuguese, on a great diversity of topics across the whole spectrum of FL.

One problem for Brazilian students is that, despite the bilingual journal, the vast majority of the essential reading is currently only available in English, although the number of relevant books and theses available in Portuguese is growing (cf. Colares 2010; Fröhlich 2014; Silveira/Abritta/Vieira 2015; Sousa-Silva/Coulthard 2016; Colares 2016; Pinto/Cabral/ Rodrigues 2016; Tullio/Gavioli-Prestes 2020; Celestino/Coulthard/Souza-Silva 2020).

An academic association for staff and students of both Linguistics and Law, and named the *Associação de Linguagem and Direito* (ALIDI) was founded in 2012. For seven years ALIDI organised country-wide seminars and conferences and was integral to the development of the discipline in Brazil, but sadly it was forced to suspend its operations in 2020 and there is currently no sign of a replacement.

In this chapter we will report research into legal language; into interaction in police stations and courts and proposals to improve current practice; into the preservation of verbal evidence; and into problems with the current provision of translation and interpreting services in police stations and courts. We will also look at the way expert services are regulated in Brazil and end by discussing some of the areas where linguists have worked as experts: plagiarism, trademark violation, defamation and an accusation of the use of racist language.

The material analysed ranges from audio recordings of interactions in legal contexts, through legal documents sent to be translated, to court judgments and even transcripts of a live interaction on YouTube. We will present and discuss some of the methodological procedures adopted and the linguistic tools employed. In order to protect the identity of the participants,

3

all names and identifying details have been removed. All the original data is in Brazilian Portuguese and although the analyses will be presented in English and illustrated with English translations, the originals will be provided.  All translations into English are our own.

## 2. The Language of the Law

Brazil is a country with strong colonialist roots, and has inherited a legal style replete with obscure language, complicated syntax, archaisms and Latinisms, labelled *juridiquês* (legalese). There have recently been some attempts to introduce into Brazil the ideas of the Plain Language movement, which started in the English-speaking world, advocating that the language used in legal-lay documents should be less complicated both grammatically and lexically to make it more accessible to the lay readers who are often the target audience. Brazilians have noted with envy that as long ago as 2010 Barack Obama signed a 'Plain Writing Act', which requires public bodies to use "clear … communication that the public can understand", and soon after, issued an Executive Order reaffirming that regulations must be accessible, consistent, written in simple language and easy to understand (Fröhlich 2014, 245).

Sadly, so far, the Plain Language movement has made little progress in the Brazilian legal community, although there is some slight evidence of movement: one Appeal Court judge, João Batista de Matos Danda, perhaps significantly from the Regional Labour Court (*Tribunal Regional do Trabalho*), deliberately used extremely simple language in one of his judgments in order to demonstrate that the judiciary can effectively communicate with the lay population (Fröhlich 2015, 234).

As a small contribution to change, we ran an experimental course for judges, examining issues of legal drafting in a 'plain language' context. Our approach was to give the judges intralingual translation tasks – that is we asked them to take barely comprehensible legal definitions and turn them into 'plain' Portuguese. This experiment was part of a Continuing Professional Development course for Judges, at the invitation of the Judicial Academy of the Court of Justice of Santa Catarina/Brazil.

**2.1 Legal Translation**

Fröhlich (2014) investigated the role of translators and interpreters in Brazil and the difficulties they encountered due to the lack of academic research, of post-graduate courses and even of basic training, not only for the translators and interpreters, but also for the legal professionals who need to use their services. We look with envy at the UK where there are courses to teach police officers how to interview through an interpreter and at Australia, where judges at the end of their training have a compulsory one-day session taught by a professor of Translation Studies considering and illustrating the legal problems of Translation and Interpreting.

Fröhlich (*ibid*) highlighted that the profession is not properly regulated in Brazil: there is no official supervising body, and not even a regulatory standard. The combination of these factors obviously directly impacts the quality of legal translation and interpreting. In this context, she proposed some specific solutions for problems involved in the translation process of key legal terms and expressions as well as various legal-linguistic habits associated with 'legalese', solutions designed to help sworn translators and interpreters to improve the quality of their translations (illustrated with reference to the translation of *rogatory letters* into and from German).

We will now address the issue of interlingual legal translation to better explain this relation. As a starting point, it is fundamental to remember that besides differences in their legal knowledge, translator/ interpreters may have varying linguistic competences. They may be a) a native bilingual; b) a native speaker of only the source language, although competent in the target language (very common in Brazil); c) a native speaker of only the target language, although competent in the source language (rarer, but preferable to option b); d) a member of a two-person team which consists of native speakers of both the source and the target languages. Brazil, as a country of continental size, presents all these aspects.

All countries have problems with suspects who do not speak the language of the legal system, and it is important to discuss the issue within the parameters of Forensic Linguistics. In Australia, for instance, police are not allowed to interview without an interpreter being present, even if the suspect claims their English is sufficiently good, so the police provide qualified interpreters free of charge. However, in Brazil, there is no police provision of qualified interpreters; instead, police officers who have some knowledge of the particular foreign

5

language may interpret, if there is no available native speaker. As a result, the evidence collected can be problematic and questionable (Fröhlich 2014).

An isolated, though promising, innovation occurred in Rio de Janeiro during the 2023 Carnival period. Due to the large number of foreign tourists in the city, bilingual guides from the Department of Tourism were sent to reinforce the tourism police. This was a very welcome step, but we have no information about the training, qualifications and indeed the competence of these tour guides, nor indeed the range of languages they could collectively cover. Sadly, the scheme did not survive the Carnival holiday period and there is no indication of whether it will be renewed next year, let alone if the scheme will be expanded to include other tourist police stations in other tourist destinations, let alone police stations more generally.  So, a very serious interpreting need remains. Several years ago, one of our masters' students spent a period of work experience at a local Tourist Police Station, where she found one enterprising officer using Google Translate with some success, although, there were predictably some infelicities.

The legal translator works at all levels of the Brazilian legal system and with a wide range of legal texts - powers of attorney, laws, statutes, contracts, wills, letters rogatory, etc. - as well as sometimes acting as an expert to evaluate third-party translations. A major problem is the need to deal with equivalence, or worse the lack of equivalence, across judicial systems and a particularly severe problem for the Brazilian translator is to render texts in *Juridiquês*, the already mentioned legal language tradition that values the use of wordy phrases and archaic terms with colonial and Latin roots, into a rhetorical tradition which prefers simple language.

To complicate matters further, a major problem in all countries, but particularly in Brazil, is the pervasive belief among lawyers and judges in the 'conduit' theory' - the belief that the meaning of the source text can be faithfully preserved, with nothing added and nothing removed, in a target language version.

## 2.2 Letters Rogatory

Considering the complexity of the areas of legal translation and interpretation in police stations and courtrooms, we have chosen for the purpose of this chapter to focus on issues involving only translation and then only in one specific legal context - the translation of *letters rogatory*.

A letter rogatory is in fact, confusingly for a lay reader, not a single document at all; the label conflates the letter itself with the appended collection of other relevant documents - these are usually petitions, powers of attorney, judgments, orders, etc. Letters rogatory are requests for legal assistance sent from the requesting court in one country to a requested court in another country with which there is a pre-existing agreement of legal reciprocity. The cases are almost always ones in which one or more of the parties involved are living outside the requesting country. In such cases, the court expert/translator is appointed by the judge. The translation process can be complicated by differences in legal conventions; these may be simple, involving only degrees of formality and terms of address, but can be much more complicated, involving differences in the subdivision of crimes into categories and/or their labelling, and specific differences in the legal vocabularies employed (Fröhlich 2014; 2015). For example, the lay Brazilian has difficulties coping with the subtle differences in meaning between the three crimes against honour: a) *Calúnia,* the false attribution of a criminal act to someone, (would this be best translated as calumny, libel or slander?); b) *Injúria*, any offense against someone's dignity, (would this be slander or libel?); and c) *Difamação* an attack on someone's reputation, (would this be defamation or slander or libel?). It will be evident that we do not know how the four English categories calumny/defamation/libel and slander fit with the three Brazilian ones and in German law the legal/language problem is even more complicated as there are at least five categories *Verunglimpfung, Verleumdung, Rufmord, Beleidigung, Diffamierung.*

A simple illustration of problems involved in legal translation is the Brazilian use of highly formal terms of address in documents. The following opening comes from a very recent rogatory letter, dated 2022: "Your Excellency, Doctor Judge of the 1st Civil Court of the District of Videira - State of Santa Catarina…" (*Excelentíssimo Senhor Doutor Juiz de Direito da 1ª Vara Cível da Comarca de Videira – Estado de Santa Catarina*). Different from Brazil, where all lawyers can be addressed as doctor (supported by the imperial law of 1827), 'Doctor' is now only used in most countries as an academic title, that is, only when a PhD degree has been awarded. In Germany, for example, there are laws that restrict the use of academic degree titles to those who have the university qualification. This raises the basic question – should legal translation be communicative, that is, should the text function as if it had been composed in the target language and in the target culture, or should it set out to give the target language reader literal access to linguistic and cultural aspects of the original message?

7

Undoubtedly, we need to consider legal translation as a widespread problem, since it affects several aspects of the legal system. In fact, there is little discussion about the real role of the translator in the Brazilian legal system. Little is said about both the specific knowledge required to perform this task, and about the translator's responsibilities, duties and rights within the judicial system. Likewise, little or nothing is said about their importance for the delivery of justice, especially when foreign parties are involved. In fact, the translator's work becomes practically invisible in the judicial system, so much so that insufficient resources are made available. There is still much that could be improved and a great deal of work for forensic linguists proselytizing and educating legal professionals.

## 3. Interaction in Legal Contexts

Police interviewing has been studied widely in the English-speaking world, and more recently research has begun in Brazil. Research focuses on the nature of the interaction between interviewers (police officers) and interviewees (lay persons); on the collection, preservation, reproduction and subsequent use in court of the resulting verbal evidence. Studies have questioned the accuracy and reliability of the evidential records and the ways in which the evidence is subsequently used in court (cf. Aldridge 2010; Haworth 2010; Rock 2001; 2010).

Brazilian researchers have admired the fact that some countries have produced guidelines for the interviewing of all witnesses, victims and suspects; and that significant interviews are audio-recorded, so that the court can have direct access to what was actually said. There are now preliminary attempts to liaise with Brazilian police to introduce international best practice, but at the moment the police have no protocols in common for obtaining and preserving verbal evidence. While the Brazilian Police Academy (*Academia Brasileira de Polícia/ACADEPOL*) does give instructions to police officers about the writing of Police reports, there is no training on how to approach interviewees in the process of data collection.

Research into police interviewing in Brazil by Jorge (2018) revealed that lack of both training and specific protocols reduces the chances of obtaining accurate evidence. For example, there are problems with the type of questions posed; police interviewers prefer closed questions, but these give the interviewer too much control over topic selection and at the same time massively reduce the likelihood that the interviewee will contribute unexpected, but useful, information.

A worrying incidental finding of this research was that many interviewers failed to deliver the warning to suspects about their right to remain silent and the information that they have a right to have a lawyer present.

In addition, there is a major problem with the preservation of the verbal evidence. A non-participating 'scribe' is present during the interview and they make a written record of (some of) what was said. There is no attempt to preserve in verbatim form the actual words the interviewee spoke, let alone what the police officer asked– much is not recorded at all and what is preserved, is usually recorded in third person past tense reported speech, so there is no subsequent access to the words the interviewee actually used - the court only has a record of what the scribe understood the interviewee to mean. Research comparing what was said with what was recorded by the scribe shows that the records consistently omit and/or misrepresent important information (Gibbons 2001; Jorge 2015; 2018). However, there is now a growing discussion about the necessity of audio and/or video recordings of police interviews.[2] Even within the police, there is a concern about the deficiencies in the way verbal evidence form interviews is recorded in written summary form. Although it is not required by law, some police departments have begun to audio- and even video-record interviews, especially interviews with suspects and offenders caught red handed. Indeed, at the time of writing a draft Federal bill is being debated, which, if passed, will require the audio/visual recording of police interviews with victims, suspects and witnesses.[3]

Interestingly, the Brazilian court system has already adopted this system, which is recognized by the judiciary as a method to provide more judicial security. The legislation says court hearings can be analogically or digitally recorded, as long as the solicitors and judges have full access to it. Also, recordings can be made by the solicitors themselves with no previous legal authorization (article 367, items 5 and 6 of the New Civil Process Code[4] and paragraph 1 article 405 of the Law 11.719/2008 of the Criminal Process Code[5]). The practice of audio and video recording has facilitated the process. Now, instead of typing, the registrar of the court simply connects the camera and the microphone to the computer. Later, these recordings are annexed

---

[2]  https://jus.com.br/artigos/89226/accountability-na-persecucao-penal-a-necessidade-de-gravar-as-provas-orais-produzidas-na-fase-extraprocessual (accessed 24 July 2023).
[3]  https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=1828280&filename=PL+5778/2019 (accessed 01 November 2019).
[4]  https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13105.htm (accessed 30 April 2023).
[5]  https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/lei/l11719.htm (accessed 30 April 2023).

9

to the digital lawsuit, which becomes available to the attorneys. In addition, as a consequence of the Covid-19 pandemic, remote court-hearings have become common practice.[6] So now, unless a case is being heard with restricted access, anyone can watch live and/or access recordings online afterwards. In this area of public justice, Brazil is well ahead of most of the rest of the world. Even police reports can now be made online (excluding serious crimes such as Rape, Homicide or robbery resulting in death).[7]

The right to remain silent, which is obligatorily communicated at the beginning of all significant British and US police interviews, is also guaranteed by law in Brazil. However, research identifies problems in its implementation. Firstly, there is no standardised official text for police officers to read out or to recite. Consequently, police officers who do inform interviewees, use a version which is usually a translation of the US Miranda Warnings. Although the warnings do indeed inform the suspect that they have the right to remain silent and/or to speak only in the presence of a defence lawyer, studies show that these rights are, not infrequently, 'waived' by the police, who simply do not issue the warnings (Jorge 2018). Secondly, warnings may be delivered, but only after the interview has already begun and the suspect has already provided pieces of (potentially incriminating) evidence (Jorge 2018). Nevertheless, there is now a growing realisation that, as is already recognised by police in the English-speaking world, this can prejudice trials. Even so, we have no examples of judges in Brazil refusing to admit evidence so collected as would happen in several countries in the English-speaking world.

On another topic, Brazil is the country with the fifth largest number of cases of violence against women and has the seventh largest number of feminicides, but comparatively low conviction rates. These facts have stimulated research by linguists into interaction in police stations and courtrooms to discover how far the practice of the legal professionals is contributing to the statistics mentioned.

Brazil has special police stations reserved for women and called *Delegacias da Mulher* (Police Stations for Women) – although interestingly, in the context of fierce debates in many European countries over the definition of 'woman', in this specific Brazilian police context *woman*

---

[6]  https://www.cnj.jus.br/aprovadas-regras-para-audiencias-judiciais-realizadas-por-meio-de-videoconferencia/ (accessed 30 April 2023).
[7]  https://delegaciavirtual.sc.gov.br/(accessed 30 April 2023).

includes two other vulnerable groups: children and old people of both sexes. However, despite this provision of dedicated police stations, studies show that cases of violence against women have in fact increased. Also, these *Delegacias* share the same interviewing problems identified above. Jorge (2018), in an analysis based on transcriptions of audio recordings of police interviews with victims, suspects and witnesses shows that the types of question put by the interviewing officers often hindered, if not actually prevented, victims from providing details of the aggression. For example, closed questions were preferred to open questions, which could have elicited a free narrative. The study also revealed that, at times, there was no coherent sequence to the questioning. Too often, police officers exercised their institutional power by dominating the interaction. On the other hand, they did show some solidarity by verbalising support for the victims. However, at the same time, through the way they coded the event(s) linguistically, the police tended to mitigate the seriousness of the aggression. Other Brazilian studies have confirmed that women's complaints tend to be treated less seriously. For example, a study conducted in four police departments in the city of Rio de Janeiro showed that women who went to report their cases tended to be viewed by male police officers as 'women who had nothing better to do at home' (Lima and Souza 2009; cf. also Nunes-Scarduelli (2017) and Jorge (2018) for more examples of police downplaying the seriousness of cases).

Domestic violence is not only downplayed by the police, but also by the judiciary. Following a study of 20 cases of domestic violence from first reporting to the conclusion of the court proceedings, Nunes-Scarduelli, (2017, 20-21), herself a serving police officer, concludes that, "[the] police officers [involved] and [the] judiciary do not share the same feelings about [the] women's situation [as do the victims], which may interfere in the verdict."[8] Nunes-Scarduelli's analysis shows that the police interviewers tend to use vocabulary choices which mitigate the violence. Also, the judiciary tends to collude with those victims who choose not to press charges, especially when they fail to turn up to court: the judicial system appears very reluctant to take into consideration that non-appearance could be the result of fear, fear that they might in future suffer revenge violence from their abusers.

The Nunes-Scarduelli study also revealed an important difference between the stance of the woman victim and the judicial system. While the women in the main reported violence not

---

8    *[...] mulheres em situação de violência, policiais e membros do poder judiciário não refletem propriamente o enfrentamento às situações de violência contra as mulheres, o que pode interferir na eficácia da aplicação desse instrumento jurídico.*

11

simply to get protection for themselves, but also to try to get treatment for their partners, because they often saw the violence as being attributable to addiction to drugs and/or alcohol– (there are instances of them insisting "my husband is a good man, but…")- the judicial system viewed the violence simply as a crime, which if proven needed to be punished by a fine or prison.

In the past ten years, other studies have focussed on the representation of women in Brazilian Criminal Courts. Canuto and Colares (2017) use CDA textual tools to identify sexist ideology in judgments in cases of rape. For example, in a case involving a 12-year-old victim, the judge observes that, although the victim's rights are protected by law, her provocative attitude should be taken into account as it was not the normal behaviour of a child. Similarly, Figueiredo (2022) discusses the discourse of a Brazilian appellate decision in a rape case in which a man had been accused of raping his female employee during a business trip. Figueiredo notes that the legal language in both the Judges' court and the Court of Appeal downplays the defendant's intention to rape. The judges noted that the defendant was not carrying a gun and also, observed that the victim had not made very much effort to reject the aggressor using as evidence that she had not suffered any physical injury.  As a result, both courts concluded that it was a case of consensual sex rather than rape. Further, Freitas and Pinheiro (2017) present an analysis of judges' narratives in cases of gender violence derived from a corpus of Brazilian appellate decisions stretching over almost a decade. The authors observed that the legal decisions relied more on the judges' perception of the level of violence in Brazil rather than on the reality.

## 4. Language as Evidence

Before starting our account of the work of the forensic linguist as expert witness in Brazil, it is important to address, if only briefly, some of the concepts and laws that regulate their work. Experts can be commissioned in two ways. Firstly, as in many jurisdictions worldwide which have investigating judges, the expert may be appointed directly by a Judge, in which case s/he will be called an Expert (*Perito*).  Secondly, they may be contracted and paid by one of the parties, but they will then be referred to as a Technical Assistant (*Assistente Técnico*). In such cases, the lawyer may choose to use the expert's report simply to inform their own

12

argumentation or they may call the expert to present the report to the court as independent evidence.

There are no specific qualifications for experts in Brazil (not even for translators and interpreters), although it is necessary for them to demonstrate technical or scientific knowledge of the area, which in the case of forensic linguists, as in many other countries, is through the possession of a relevant academic degree, and/or academic publications and/or previous experience in the area. Experts can register their availability with various entities, such as the State Courts, the Public Prosecutor's Office and the Order of Brazilian Attorneys (*Ordem dos Advogados Brasileiros/OAB*), all of which maintain a register of experts for legal professionals to consult.   However, in order to register, the expert does not necessarily need academic qualifications in their area of expertise, so appearance on the list is not an absolute guarantee of technical/scientific expertise.

We will now present accounts of some Brazilian cases and link them, where useful, to earlier cases in other countries. All the Brazilian data was obtained in confidence and so, although extracts will be used to exemplify the cases we discuss, the identity of the participants will not be revealed.

As far as we know, the first time a forensic linguist was commissioned in Brazil was for a case of suspected academic plagiarism where a PhD student, about to submit her dissertation, accused a recently graduated colleague of having plagiarised three pages of her conclusion. An early version of the program *Copycatch* (Woolls/Coulthard 1998) was used and the results demonstrated that, in fact, several chapters of the thesis contained unacknowledged borrowings. The award was cancelled and the newly appointed lecturer lost his job.


### 4.1 Whose meaning?

Many cases concern disputes about meaning, the meaning of individual words, of phrases, even occasionally of whole sentences and sometimes there are disagreements about what illocutionary or perlocutionary act has been performed. There immediately arises the question of *whose meaning*; that is, whose interpretation will be used as the basis of the judgment; is it what the speaker/writer says they intended, what the aggrieved party claims they understood,

13

what the judge/jury decide or what a reasonable uninvolved person would be thought to have understood. It can be even more complicated if the disputed item is in a foreign language. In 2020 the Uruguayan footballer, Edinson Cavani, who at the time was playing in the English Premier League, replied in Spanish to a congratulatory Instagram post from a Uruguayan fan, with the phrase *Gracias negrito*. The phrase was translated literally by the English Football Association as 'thanks little black person', and was interpreted as a racist comment. As a direct consequence, Cavani was suspended for three games and fined £100,000 (worth some $132,000 at the time) for adjudged racist behaviour. Cavani protested that the Spanish word *negrito* was intended, and would have been so understood by the recipient, and indeed by any Uruguayan, as an expression of affection. *Negrito*, although being derived from a lexical root 'negr*', which does have as one of its meanings 'black', has no colour meaning in this usage. Indeed, the son of one of the authors, when expressing affection to his wife in Portuguese, calls her *pretinha*, literally 'little black female', even though she is tall and of Italian descent – the word has no residual skin colour meaning in this usage and is certainly not understood as a racist comment by the wife. However, similarly, none of the contextually intended affective meaning of these two terms of 'endearment' is conveyed by literal translation into English. Cavani was supported in his interpretation not only by fellow Uruguayan footballers, but also by the Uruguayan Academy of Letters, which denounced the imposed sanction as an example of English football's lack of 'cultural and linguistic knowledge'. Nevertheless, the penalty was not rescinded and the 'guilty of racism' verdict stood.

Shortly afterwards we had a similar case in Brazil. A black footballer who has a large quantity of curly hair claimed that he had heard a member of the opposing team's technical staff shout a racial insult at him: *Vai cortar esse cabelo de cachopa de abelha,* ("Go cut your beehive hair"). The defence countered that the expression had not actually been uttered – indeed a forensic phonetician submitted a technical report, with acoustic and word-level analysis of an audio recordings of the match, which asserted that at no time were there any offensive phrases uttered and certainly not any word similar to *Cachopa*. But in any case, the defence continued, even if the phrase had been uttered, while offensive, the utterance was not racist.

We examined the use of the term *Cachopa* to refer to other sportsmen. In fact, there are several Brazilian athletes who are known as *Cachopa* and the use is not seen to be negative and certainly not racist. Nicknames, which many non-Brazilians would regard as insulting, are quite frequent in Brazil. English football commentators, for example, happily referred to a previous

14

manager of the Brazilian national football team as *Dunga,* assuming it was his surname, without ever realising it was a nickname and the Portuguese name for the seventh dwarf in Snow White and the Seven Dwarfs.  So they were, like everyone in Brazil, calling him *Dopey*.

Thus, with no context or surrounding linguistic co-text, it is hard to argue for a racist meaning. However, in Brazilian interaction any reference to the hair of black people is now seen as essentially racist, even if the words used are not themselves lexically racist. Given this context, the footballer's claim was upheld and the opposing team was fined, in accordance with the legal protection against racism in Brazil (Law nº 7.716/89 and Law 14.532/2023).

## 4.2 Cases of Suspected Plagiarism

Plagiarism happens when someone borrows another's text and uses it as their own without acknowledging the original author. In Brazil, plagiarism is a crime protected under Law n° 9.610, [9] which describes the author's right to ownership. Nowadays, to help detecting plagiarism, there are computer programs such as *Copycatch* and *Turnitin*, the latter used widely in academic institutions for checking student essays. With *Turnitin* and similar Brazilian programs in common use, no one now needs an expert of Forensic Linguistics to detect plagiarism, although there will always be questions from lawyers and judges about how much similarity can occur by chance (Love 2002).

Coulthard (2004) argued that anyone generating a new meaningful, grammatically organised, sequence of words quickly produces a sequence that is unique, that is, they generate a sequence that no one, not even the author themself has ever produced before. Further empirical research over 20 years allows us to say that if two texts share an identical unattributed 10-word sequence, for which Google can find no other instances, there is a possibility that one author copied from the other or that both copied from a third text. If the identical sequence is even longer and especially if there are also several other such shared long identical sequences, and/or other almost identical shorter sequences linked by deletion, addition or paraphrase, the evidence for plagiarism can become persuasive.

---

[9]  https://www.planalto.gov.br/ccivil_03/leis/l9610.htm (accessed 30 April 2023).

15

However, it can be more difficult to correctly attribute significance to the identified plagiarism. The poet T.S. Eliot reported:

> "In one of my early poems ['Cousin Nancy'] I used, without quotation marks, the line 'the army of unalterable law...' from a poem by George Meredith, and this critic accused me of having deliberately plagiarized, pinched, pilfered that line. Whereas, of course, the whole point was that the reader should recognize where it came from and contrast it with the spirit and meaning of my own poem (quoted in Ricks 1998, 23–24)."

Growingly, the detection of plagiarism is not an end in itself for forensic linguists. So, for example, in a famous UK case, known as the 'Bridgewater Four', where four men were convicted of murdering, a 14-year-old boy, mainly on the basis of a disputed confession, the conviction was quashed after the existence of plagiarised text in a contested police interview record was used to demonstrate that the record had been falsified by copying several utterances verbatim from an unchallenged statement (Coulthard/Johnson 2007, 191-196).

We have recently been involved in several cases where it has been claimed that judges have committed plagiarism in producing their judgments. With regard to textual borrowing in legal texts, there is much to talk about. Legal Judgments, for example, are complex documents and their composition can involve several separate author roles. Firstly, there is the executive author; the person who drafts the text; then there is the precursory author, an author whose text is incorporated by the executive into their text. In the case of a judgment the precursory text may be quotations from statutes or from previous judgments as well as argument and evidence that was presented during the trial. Precursory text may be marked in judgments by various devices: quotation marks, italics, indentation as well as by linked footnotes indicating sources. Finally, there is the declarative author who may or may not have written (all/some/any of) the text, but who takes responsibility for it. So, for example, it is not unknown for a judge to sign a judgement which has been entirely written by assistants.

In one case we worked on in Brazil, a defence lawyer had noticed what she thought was a plagiarised sentence in the Judgment, that is text that she thought, had been borrowed without acknowledgement from the Public Prosecutors' written Argument. In other words, the lawyer was claiming that precursory text derived from the Argument had been presented as if it had been created by the judge himself as executive author. In fact, we were able to show that there was very much more plagiarised text than the lawyer had suspected. The judge had divided his 145-page judgment into three sections, with the middle section containing 64 pages. In this central section, only two relatively short extracts had been marked as quoted from the

16

prosecutors' argument that is marked as precursory text; however, analysis showed that 51 of these 64 pages contained unacknowledged borrowings. Indeed, several pages consisted almost exclusively of plagiarised text.

After more analysis, it became evident that these plagiarised extracts had not even been retyped from the prosecution document; instead, the executive author/judge had simply edited an electronic version of the Argument, and so his version also shared spelling and grammatical mistakes with the source text. In so copying, the judge had presented, as if they were his own, not only many of the prosecution arguments, but also, as supportive evidence, extracts which they, not he, had selected from transcripts of audio-recorded interviews with the accused.

The obvious question was why? Why did the judge choose to plagiarise and thereby lead the reader to assume that the arguments and supporting quotations were his own, rather than present the borrowed precursory prosecution text in quotation marks, and then demonstrate why he agreed with their arguments. Obviously, one would expect a judge to interact with and then evaluate the evidence presented to him, but in this case, he had on the one hand simply incorporated the prosecution arguments and supporting evidence with no overt evaluation, while on the other he had not used a single quotation from the defence's substantial pleading document, not even to indicate why he disagreed with it.

The judge has not responded to our findings, but the evidence at least allowed the Defence to question the neutrality of a judge who on the one hand conceals his debt to the Prosecution and on the other makes no attempt to engage with the arguments advanced by the Defence. Our findings were a substantial part of the Appeal documents and at the time of writing we have just heard that the conviction has been quashed.

In another Brazilian case the defendant claimed that the judgment had been partially or even totally written by someone other than the judge. Up to page 8, with a few exceptions, the introduction was plagiarized practically *ipsis litteris* from the prosecutors' argument.  As a consequence, rather than being a simple factual account of the main events in the case, on which the Judge would later give his commentary, the account incorporated many of the prosecutors' denunciatory words and phrases and so prejudged the conclusion. There is, for example, a reference in the introductory account of the facts of the case to a 'corrupt police officer', who was thereby linguistically condemned before the judge began to articulate his opinion. And in this case, it was possible to go even further, because we had three other

17

Judgments authored by the same judge and so it was possible to show that there were significant stylistic differences between the disputed judgment and the other three. Finally, there were quite a lot of linguistic and paralinguistic features in the document which we argued could not have been produced by a judge versed in legal writing – for instance numerous spelling, grammatical and stylistic errors. We do not know whether the judge accepted the taint of professional linguistic incompetence or preferred to admit plagiarism.

Brazil has a long history of academic plagiarism by translation too – not only by students but also by staff – and as the free translation programs – Google Translate, Deepl, ChatGPT and other tools – get better and better such plagiarism becomes harder and harder to detect, although most plagiarists do not realise that, if the text is submitted shortly after the translation, a suspicious reader can re-translate the text and then search for the original on line (cf. Sousa-Silva 2017).

### 4.3 Trademark Violation

Trademark violation has some similarities to plagiarism, in that it is an attempt to benefit from the work of others, but it differs radically in that the whole purpose is for the textual link to be obvious in order to bask in reflected glory, so it is better seen as a case of *intertextuality*. The Brazilian national drink, distilled from cane sugar, is *cachaça* and there are literally hundreds of different brands for sale, but whisky is the distilled drink with the highest prestige in Brazil and the whisky consumed most is *Johnny Walker Red Label*. Several years ago, a producer launched a *cachaça* labelled *João Andante*,[10] (literally *John the Walker*) which had on its label an image of a famous historical figure of the same name, a 'wanderer' whose was typically represented, like the Englishman Dick Whittington, with a bag of possessions supported on a stick over his shoulder.

The intertextual link was transparent and it was seen as a humorous gesture linking a new brand of *cachaça* to a well-established whisky brand. However, Diageo, producer of *Johnnie Walker* whisky, sued claiming that their brand had been appropriated and that, despite *cachaça* being a totally different drink and that Diageo did not even sell *cachaça*, the brand was trying to benefit

---

[10]    Images can be seen at https://zenitemarcas.com.br/cachaca-joao-andante-vs-johnnie-walker/ (accessed 30 April 2023).

18

from *Johnnie Walker*'s hard-won reputation for quality. The prohibition on using a registered trademark stems from Law 9.279/1996, which imposes on everyone the duty to respect the right to exclusive use of the registered trademark. A well-known trademark in its field of activity under the terms of article 6 bis (I) of the Paris Union Convention for the Protection of Industrial Property enjoys special protection, regardless of whether it has been previously deposited or registered in Brazil.[11] Despite the respondents' claim that they were not using 'the registered trademark', that the differences in the product, the name, the colours of the label and the image, Diageo won. In response, the *cachaça* manufacturer adapted its original whimsical labelling, by changing the name of the *cachaça* to the still whimsical *O Andante* (*The Walker*) and removing the legs from the image of the walker, so he became a legless walker.

Interestingly, another manufacturer then produced a female version, called *Maria Andante*, (*Mary the Walker*) with the image of a non-historical woman (and with legs) on their label. Diageo sued again, but it lost in the second trial, although we do not have access to the judges' reasoning. These cases had such an impact that they generated numerous comments in various media, including being referred to in a master's thesis submitted to the National Institute of Industrial Property,[12] and bottles with the original label sold at a premium.

### 4.4 Language as Evidence in a Disparagement Case

We will end this section with a recent case which illustrates the application of several important insights to the same set of data. The initial narrative is quite simple: a young Brazilian sportsman signed a contract with the American owners of a professional team, which included a 'non-disparagement' clause. When his contract ended, he signed a second contract which also included a 'non-disparagement' clause. Months later, while he was in Brazil, he took part in a 'live', an online Brazilian sports-focussed chat programme, during which everyone spoke Portuguese and comments were made about the teams he used to play for and some of the problems players had experienced. Several days later lawyers contracted by the American owners submitted an indictment that alleged that, during the live, he had disparaged the

---

[11] http://www.planalto.gov.br/ccivil_03/leis/l9279.htm (accessed 30 April 2023).
[12] https://www.gov.br/inpi/pt-br/servicos/a-academia/arquivo/dissertacoes/SILVAMariaClaradeOliveira.pdf (accessed 30 April 2023).

company at least seven times and they sued for compensation at a rate of $100,000 per disparagement.

The lawyers, as well as the team owner, were Americans based in the USA and none of them claimed to speak Portuguese, so the evidence they cited in support of their claim consisted of English translations of decontextualised Portuguese utterances – in fact only 85 words out of some 25,000 uttered were actually cited. The lawyers proposed going to arbitration, with an American arbitrator in Los Angeles, where it was highly probable that, apart from the defence lawyers, none of the legal professionals attending the tribunal would understand the Portuguese evidence and so would work exclusively from the very partial and poor-quality written translations.

We were asked to write a report to help the defence. The main linguistic points we made were:

- The contractual system was linguistically unfair, because the contract was only presented in English to a non-native speaker and there was apparently no attempt to check that he understood it. Crucially, for the purpose of the case, the meaning of 'disparagement', which was used as a 'term of art', was crucial, but it was not defined, let alone exemplified, neither in the contract, nor even in the indictment;

- The termination contract only committed the sportsman and not the company to not disparage the other;

- The significant wording of the crucial contract clause was "shall not **intentionally** (our highlighting) disparage the reputation of…". We pointed out: a) that it is not possible to access a speaker's intention, so, the chosen wording was actually unenforceable; and b) that 'disparaging' is not an illocutionary act in English - that is, one cannot say "I hereby, disparage you…" - rather, disparaging is a perlocutionary act achieved as a consequence of a different illocutionary act, and so, as speakers have no control over perlocutionary acts, they literally cannot 'intentionally disparage';

- Even had the word 'intentionally' not been included in the clause the Arbitrator's task would not have been easy. The job of the Tribunal would then have been to rule on whether the company had been disparaged, but the Arbitrator would first need to decide which of several potentially distinct meanings of the quoted extracts to choose. Would it be the meaning that the speaker (the sportsman) says he intended to convey; or the

20

meaning the addressees (those participating with him in the live) said they understood to have been conveyed; or the meaning the listeners (members of the non-participating audience) said they understood; or the meaning the overhearers (representatives of the company attempting to access the meaning through English translations of extracts from the live) said they understood; or the meaning that the Arbitrator judged would have been understood by a reasonable unbiased third party; or perhaps even the Arbitrator's own understanding;

- The evidence was spoken Brazilian Portuguese with lots of slang and obscenities, but the evidence on which the lawyers relied was an English translation of a written transcript. We also argued, unsuccessfully, that the case required an Arbitrator who, if not Brazilian, at least spoke Brazilian Portuguese;

- The situation was further complicated because, as in so many cases worldwide, the lawyers assumed that a written transcription of what was said was equivalent to the original. We pointed out that Haworth (2018), as already discussed above, has shown the many ways in which transcribed evidence can be unfaithful to the original and misleading, while noting that recent research has shown that jury members working with written evidence have a lower opinion of the reliability and truthfulness of a given witness than do those who listen to a recording of their evidence;

- In this case, there was the added complication of spoken Portuguese evidence being presented in written English form and treated as if the accused had spoken English. We pointed out that, as already noted above, most lawyers, unaware of the aphorism *traduttore/traditore,* (translator/traitor), naively assume that a translation is communicatively equivalent to the original. We were helped in making our point about the unreliability of the translated evidence, when it turned out that the translations cited in the original indictment had actually been made by a third-year undergraduate student of medicine rather than, as required in cases that are tried in Brazilian courts, by a certified translator;

- We reinforced our warning about the unreliable quality of the translations by citing two examples: the sportsman's observation 'o time é foda', literally 'the team is fuck', had crucially been mistranslated as a criticism, 'the team is shit', rather than as the intended 'the team is wonderful/amazing', because the translator had failed to spot that this was a classic example of reverse meaning, and was intended to be positively evaluative.

21

Secondly a phrase 'meio lunático', (literally 'half mad'), when applied to one of the admired team owners, was translated as 'a little lunatic' instead of as the communicatively faithful 'a little bit crazy'. We were able to use these two examples, along with others, as evidence that individual meanings could only be accurately decoded inside a much larger linguistic context, which those not fluent in Portuguese could not decode;

- Finally, we argued that, just as, to fully appreciate it, one must watch a foreign film with the original soundtrack and subtitles, so it was essential for the linguistic evidence in this case:

    a) to be presented to the court in its original spoken form, so that the arbitrator would have access to the original voice quality and would be able to assess any paralinguistic evidence of aggression, playfulness etc.;

    b) to be presented to the court in its wider linguistic context – we showed how locating a claimed disparaging 7-word phrase within its 200 running-word surrounding context demonstrated that the decontextualised meaning asserted by the plaintiff's lawyers was the exact opposite of the phrase's contextualised meaning; and;

    c) to be presented to the court in a subtitled version using pre-agreed translations.

Sadly, the parties settled before the case went to court, so we do not know what the Arbitrator would have made of our arguments, although the defence lawyers said that our report was crucial in achieving what they felt was a good financial settlement for their client.

## 5. Final Remarks

This chapter has provided an overview of the achievements of Forensic Linguistics in Brazil, both academically and professionally. There have been academic events and short courses in prestigious Universities, which have generated interest in the area among legal scholars and professionals. There has been some significant research on Brazilian data and published in Portuguese, a bilingual journal has been established and PhD dissertations have investigated interaction in police stations and courts, particularly in cases of violence against women. And

also, linguists have been involved as experts in cases of disputed authorship, plagiarism disparagement and translation.

Even so legal professionals have been slow to welcome help from forensic linguists. This is partly because many legal professionals are unaware of what linguists can contribute: while on the one hand, many legal professionals do not consider the work of linguists to be relevant, many others are hindered by a lack of competence in English because much of the pertinent literature in not yet available in Portuguese.

So, while forensic linguists emphasise the intersection between the two academic areas, Linguistics and Law, it is crucial to continue to offer specialized courses to lawyers and to expand the number of publications in Portuguese so that Brazilian Legal professionals can become better informed about the work of linguists and the ways in which it can contribute to solving legal problems.

# 6. References

Aldridge, Michelle (2010), *Vulnerable witnesses in the criminal justice system*, in: Malcolm Coulthard/Alison Johnson (eds.), *The Routledge Handbook of Forensic Linguistics,* London, Routledge, 296-314.

Canuto, Sheyla/Colares, Virgínia (2017), *A representação da mulher no sistema jurídico penal: um estudo de caso a partir das análises das expressões referenciais,* Language and Law / Linguagem e Direito 4(2), 72–88, retrieved from https://ojs.letras.up.pt/index.php/LLLD/article/view/3284 (21.04.2023).

Celestino de Almeida, Dayane/Coulthard, Malcolm/Sousa-Silva, Rui (eds.) (2020), *Perspectivas em Linguística Forense,* Campinas, IEL/UNICAMP.

Colares, Virgínia (ed.) (2010), *Linguagem e Direito*, Recife, Editora Universitária (UFPE).

Colares, Virgínia (ed.) (2016), *Linguagem e Direito: caminhos para linguística forense*, São Paulo, Cortez.

Coulthard, Malcolm (2004), *Author identification, idiolect and linguistic uniqueness,* Applied Linguistics, 25 (4), 431-447. DOI: https://doi.org/10.1093/applin/25.4.431.

Coulthard, Malcolm/Johnson, Alison (2007), *An Introduction to Forensic Linguistics: Language in Evidence*, London, Routledge.

Figueiredo, Débora (2022), *The discursive dispute around the meanings of 'rape', 'consent' and 'violence' in a Brazilian appellate decision*, Revista Latinoamericana de Estudios del Discurso 22(1), 24–36. DOI: https://doi.org/10.35956/v.22.n1.2022.p.24-36 (21.04.2023).

Freitas, Lúcia/Pinheiro, Veralúcia (2017), *Narrativas de violência de gênero em acórdãos do STJ sobre Lei Maria da Penha,* Language and Law / Linguagem & Direito 4(2), 36–49. DOI: https://ojs.letras.up.pt/index.php/LLLD/article/view/3282 (21.04.2023).

Fröhlich, Luciane (2014), *Tradução Forense: Um Estudo de Cartas Rogatórias e suas Implicações* (Doctoral thesis), Federal University of Santa Catarina.

Fröhlich, Luciane (2015), *Redação jurídica objetiva: o juridiquês no banco dos réus,* Revista da Escola Superior da Magistratura do Estado de Santa Catarina (ESMESC) 22 (28), 211-236. DOI: https://doi.org/10.14295/revistadaesmesc.v22i28.p211.

Gibbons, John (2001), *Legal transformations in Spanish: An 'audiencia' in Chile*. Forensic Linguistics 8 (1), 24–43.

Haworth, Kate (2010), *Police interviews in the judicial process,* in: Coulthard, Malcolm/May, Alison (eds.), *The Routledge Handbook of Forensic Linguistics*, London, Routledge, 169-181.

Haworth, Kate (2018), *Tapes, transcripts and trials: the routine contamination of police interview  evidence*. The International Journal of Evidence and Proof 22(4), 428-450. DOI: https://doi.org/10.1177/1365712718798656.

Jorge, Sabrina Silveira de Souza (2015), *Coerência local e global em textos de relatos de ocorrências criminais*. In: Coulthard, Malcolm/Colares, Virgínia/Sousa-Silva, Rui (eds.), *Linguagem e Direito: Os Eixos Temáticos*. Recife, ALIDI, 1, 279 – 296, retrieved from https://pt.scribd.com/document/373541676/E-Book-Linguagem-e-Direito-Os-Eixos-Tematicos# (23.4.2023).

Jorge, Sabrina Silveira de Souza (2018), *Analysing Brazilian Police Interviews in Cases of Violence against Women* (Doctoral thesis), Federal University of Santa Catarina.

Lima, Lana Lage da Gama/Souza, Suellen André de (2009), *Representações de gênero e atendimento policial a mulheres vítimas de violência*, Revista Internacional Interdisciplinar INTERthesis 6 (2), 61-85. DOI: https://doi.org/10.5007/1807-1384.2009v6n2p61 (22.04.2023).

Love, Harold (2002), *Attributing authorship: An introduction,* Cambridge University Press.

Nunes-Scarduelli, Márcia (2017), *Mulheres em situação de violência conjugal*, Language and Law/ Linguagem e Direito 4(2), 19-35, retrieved from https://ojs.letras.up.pt/index.php/LLLD/article/view/3281 (22.04.2023).

Pinto, Rosalice/Cabral, Ana Lúcia Tinoco/Rodrigues, Maria das Graças Soares (eds.) (2016), *Linguagem e Direito: perspectivas teóricas e práticas*, São Paulo, Contexto.

Ricks, Christopher (1998), *Plagiarism*, in: Proceedings of the British Academy 97/ 1997 Lectures and Memoirs. 149-168.

Rock, Frances (2001), *The genesis of a witness statement*, Forensic Linguistics, 8, 44-72.

Rock, Frances (2010), *Witnesses and suspects in interviews. collecting oral evidence: the police, the public and the written word*, in: Coulthard, Malcolm/May, Alison (eds.), *The Routledge Handbook of Forensic Linguistics*, London, Routledge, 154-166.

Silveira, Sonia Bittencourt/Abritta, Carolina Scali/Vieira, Amitza Torres (eds.) (2015), *Linguistica Aplicada em Contextos Legais,* São Paulo, Paco Editorial.

Sousa-Silva, Rui. (2017), *Detecting translingual plagiarism and the backlash against translation plagiarists. Language and Law / Linguagem e Direito*, 1 (1), 70-94, retrieved from https://ojs.letras.up.pt/index.php/LLLD/article/view/2444 (22.04.2023).

Sousa-Silva, Rui/Coulthard, Malcolm (2016), *Linguística Forense,* in: Dinis-Oliveira, Ricardo/Magalhães, Teresa (eds.), *Ciências Forenses*, Lidel, Porto, 137-44.

Tullio, Cláudia/Gavioli-Prestes, Cindy (edd.) (2020), *Linguística Forense: reflexões e debates,* Ponta Grossa, Texto e Contexto, retrieved from https://www.textoecontextoeditora.com.br/assets/uploads/arquivo/a02f3-ebook-linguistica-forense.pdf. (21.07.2023).

Woolls, David/Coulthard, Malcolm (1998), *Tools for the Trade*. Forensic Linguistics 5 (1), 33-57.

DR. MALCOLM COULTHARD
m.coulthard@aston.ac.uk

DR. LUCIANE FRÖHLICH
lu.frohlich@gmail.com

DR. SABRINA JORGE
sabrinajorge@hotmail.com

25

**<u>CERTIFICATE OF SERVICE</u>**

I hereby certify that on August 31, 2023, I caused a true and correct copy of the foregoing

Defendants Tracy Deebs-Elkenaney p/k/a Tracy Wolff, Entangled Publishing, LLC, Holtzbrinck

Publishers, LLC d/b/a MacMillan, and Universal City Studios LLC's Notice of Rebuttal Expert

Disclosure and accompanying Expert Rebuttal Report by Dr. Malcolm Coulthard to be served on

the following:

Mark D. Passin, Esq.
CSReeder, PC
11766 Wilshire Blvd, Suite 1470
Los Angeles, CA 90025
(310) 861-2475
mark@csrlawyers.com

Stephen M. Doniger, Esq.
London L. Zamora, Esq.
DONIGER / BURROUGHS
247 Water Street, First Floor
New York, New York 10038
(310) 590-1820
stephen@donigerlawfirm.com
Attorneys for Plaintiff

Lacy Herman Koonce, III
Klaris Law PLLC
29 Little West 12th St.
New York, New York 10014
(718) 974-8717
lance.koonce@klarislaw.com
Attorneys for Defendants
Emily Sylvan Kim and Prospect Agency LLC

/s/ CeCe M. Cole
CeCe M. Cole